DOCUMENT RESUME

ED 342 801 TM 017 968

AUTHOR	Aman, Michael G.
TITLE	Assessing Psychopathology and Behavior Problems in Persons with Mental Retardation: A Review of Available Instruments.
INSTITUTION	National Inst. on Alcohol Abuse and Alcoholism (DHHS), Rockville, Md.
REPORT NO	DHHS-ADM-91-1712
PUB DATE	91
NOTE	241p.
PUB TYPE	Reference Materials - Bibliographies (131) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC10 Plus Postage.
DESCRIPTORS	*Behavior Disorders; Developmental Disabilities; *Diagnostic Tests; Evaluation Methods; Interviews; *Measures (Individuals); Medical Evaluation; Mental Disorders; Mental Health; *Mental Retardation; *Psychological Testing; *Psychopathology; Surveys

ABSTRACT

This survey of the instruments and methods that are currently available for assessing mental health problems in persons with mental retardation lists formalized instruments and interview techniques and evaluates them from a methodological perspective. Emphasis is on the assessment and classification of disorders rather than on the evaluation of adaptive behaviors or treatment effects. Information was solicited from several professional organizations with an interest in behavior, psychopathology, and developmental disabilities through letters sent to 50 prominent researchers and through computer searches of the literature. Approximately 40 relevant instruments were identified. These are described in three sections: (1) the more established instruments, most of which have been published, with detailed descriptions and thorough critiques; (2) relatively new or unpublished instruments, with brief summaries and critiques; and (3) relevant instruments considered peripheral to assessment of behavior disorders, with brief descriptions and no appraisal of psychometric characteristics. Eight tables summarize information about the instruments. Three appendices provide supplemental information about the survey process and the instruments reviewed. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED342801

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

National Institute of Mental Health

ASSESSING PSYCHOPATHOLOGY AND BEHAVIOR PROBLEMS IN PERSONS WITH MENTAL RETARDATION:

A REVIEW OF AVAILABLE INSTRUMENTS

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Alcohol, Drug Abuse, and Mental Health Administration

TM017968

BEST COPY AVAILABLE

ASSESSING PSYCHOPATHOLOGY AND BEHAVIOR PROBLEMS IN PERSONS WITH MENTAL RETARDATION: A REVIEW OF AVAILABLE INSTRUMENTS

Report Prepared for the National Institute
of Mental Health

Michael G. Aman
The Nisonger Center for
Mental Retardation and Developmental Disabilities
The Ohio State University

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Alcohol, Drug Abuse, and Mental Health Administration

National Institute of Mental Health
5600 Fishers Lane
Rockville, Maryland 20857

Acknowledgements

This review was funded by a Professional Services Contract from the National Institute of Mental Health (NIMH) to Michael Aman, Ph.D. The author would like to acknowledge the important role of Dr. Eleanor Dibble of the NIMH who first envisaged the need for such a review, who initiated the funding of the project, and who provided many useful suggestions on how the review might be structured. The author would like to thank the following workers for professional advice and critical comments on aspects of this research: Drs. Fred Damarin, David Hammer, Johannes Rojahn, Donald Smeltzer (Ohio State University), Dr. Stephen Schroeder (University of Kansas), and Dr. Nirbhay Singh (Medical College of Virginia). The author is grateful to Peggy Smith (Virginia Treatment Center for Children) for her many useful editorial suggestions and, as always, to Marsha Aman for a great deal of moral support and practical help including substantial editorial advice and word processing of this document.

Suggested citation: Aman, M.G. (1991). *Assessing psychopathology and behavior problems in persons with mental retardation: A review of available instruments.* Rockville MD: U.S. Department of Health and Human Services.

Table of Contents

Acknowledgements iii

Introduction 1

Survey Methods Employed 3

Selection Criteria for Instruments 4

Instruments Not Included 5

Other Instruments Relevant to the Assessment of Psychopathology 7

Evaluation Criteria 8

Acceptable Ranges 9

Review Format 10

Some Caveats 10

The Nature of Psychopathology in Mental Retardation 12

Part I 21

AAMD Adaptive Behavior Scale: Residential and Community Edition 23

AAMD Adaptive Behavior Scale: School Edition 30

Aberrant Behavior Checklist 35

Adolescent Behavior Checklist 41

Balthazar Scales of Adaptive Behavior: II. Scales of Social Adaptation 46

Behaviour Disturbance Scale 51

Client Development Evaluation Report 55

Clinical Interview Schedule 59

Devereux Adolescent Behavior Rating Scale 65

Devereux Child Behavior Rating Scale 70

Diagnostic Assessment of the Severely Handicapped (DASH) 75

Emotional Disorders Rating Scale: Developmental Disabilities 80

Minnesota Developmental Programming System (MDPS): Behavior Management Assessment 84

Preschool Behavior Questionnaire 89

Prout-Strohmer Personality Inventory 93

Psychopathology Instrument for Mentally Retarded Adults (PIMRA) 97

Reiss Screen for Maladaptive Behavior 103

Schedule of Handicaps, Behaviour, and Skills (HBS) – Revised 107

Self-Report Depression Questionnaire 113

Strohmer-Prout Behavior Rating Scale 117

Vineland Adaptive Behavior Scale 121

Part II 125

Attention Checklist 127

Behavior Development Survey 128

Behavior Evaluation Rating Scale (BeERS) 130

Behavior Inventory for Rating Development (BIRD) 131

Behavior Problems Inventory 133

Communication Style Questionnaire 135

Developmentally Delayed Children's Behaviour Checklist 137

Fairview Maladaptive Behavior Survey 139

Gilson-Levitas Diagnostic Criteria 141

Motivation Assessment Scale 143

Psychosocial Behaviour Scale 145

Revised Children's Manifest Anxiety Scale 147

Social and Emotional Behavior Inventory 150

Social Judgment Scale 152

Social Participation Rating Scale 154

Standardized Assessment of Personality 156

Structured Clinical Interview 158

Zung Self-Rating Anxiety Scale (Adapted) 159

Part III 161

Cognitive Diagnostic Battery 163

Maladaptive Behavior Scale (MABS) 164

Paroxysmal Behavior Scale 165

Seizure and Related Behavior Checklist 166

Shortened Stockton Rating Scale 167

Social Performance Survey Schedule 168

Vocational Problem Behavior Inventory 169

Conclusions and Recommendations 171

Characterization of Existing Scales 173

State of the Field 174

Recurring Problems with Available Instruments 178

Toward a Valid Taxonomy of Emotional and Behavior Disorders in Mental Retardation 179

Recommended Instruments 181

Recommendations for Future Research 183

Tables 187

Table1: Magnitude of Interrater Reliability Correlations for Different Types of Informant 187

Table 2: Psychometric Features Addressed in Reviewed Instruments (Part I) 188

Table 3: Instruments in Parts I and II Classified by Type of Rater 190

Table 4: Instruments in Parts I and II Classified by Age Group Covered 191

Table 5: Instruments in Parts I and II Classified by Method of Derivation 192

Table 6: Instruments in Parts I and II Classified by Number of Subscales 193

Table 7: Instruments in Parts I and II Classified by Level of Mental Retardation Covered 194

Table 8: Commonalities among Factors from Factor Analytic Research 195

Appendices 197

Appendix A: Societies and Associations Whose Memberships Were Notified Regarding the Review 197

Appendix B: Summary of Psychometric Characteristics of Reviewed Scales 198

Appendix C: Full Instrument Names for Abbreviations Used in Tables 3 Through 7 238

Introduction

In recent years, there has been great interest, both in the United States and in other countries, in the nature and appropriate methods for assessing mental health problems in persons with mental retardation. This has led to a number of activities such as the following. In May 1986 the National Institute of Mental Health (NIMH) convened a special workshop on the topic of "Methodological problems in treatment research with mentally retarded populations who are also mentally ill" (see Special Feature on Treatment Research, 1986). A second NIMH-sponsored workshop was held in February 1987 on "Assessment and treatment of psychiatric disorders in mental retardation." In addition, related presentations were made during 1986 and 1987 in national meetings of the National Association of the Dually Diagnosed, The American Association for Mental Retardation, and an International Research Conference on Mental Health Aspects of Mental Retardation (see Reiss, 1989). An opinion that emerged repeatedly at many of these workshops and conferences was that a lack of uniform or adequate assessment instruments has hampered clinical research. Many studies have employed idiosyncratic or individualized methods of assessment, and this has hindered comparison across investigations. However, it was not clear how accurate this impression was of the actual need for better diagnostic instruments. Thus, there appeared to be a considerable need for a systematic survey of the instruments and methods that are currently available for assessing mental health problems in persons with mental retardation.

The present project was carried out to help meet this requirement. One objective was to collect all formalized instruments and interview techniques for evaluating psychopathology and behavior disorders in persons with mental retardation. The second principal objective was to describe these instruments and to evaluate them from a methodological perspective. It is hoped that this will help to inform interested workers about the available pool of assessment techniques and their relative merits. It should be noted that the emphasis in this project has been on assessment and classification of disorders *per se* rather than on the evaluation of adaptive behavior or treatment effects. Thus, instruments developed to measure adaptive behavior or treatment effects could come under the terms of this review, but the evaluation necessarily was directed to diagnostic precision.

Survey Methods Employed

A variety of methods was used to identify and locate appropriate rating and diagnostic instruments. Extensive efforts were made to inform workers in the field that the assessment was underway and to seek submissions of all relevant materials, whether published or not. These efforts included the following:

1. Notices were sent to a number of societies and organizations whose membership was known to have an interest in behavior problems, psychopathology, and developmental disabilities. In each case, a notice described the objectives of the review project and asked that all relevant materials be sent to the author. The organizations that were contacted are listed in Appendix A.
2. Computer searches were conducted to examine the literature for relevant publications on the assessment of behavior problems and/or dual diagnosis. These included Medline, BRS (Psych Info), and BRS Health Instruments File Database searches.
3. Personal letters were sent to 50 prominent researchers who were known to be interested in assessment research in the mental retardation field. This was expedited by the literature search discussed above and by suggestions provided by colleagues in the field. The individuals who were contacted resided in eight different regions including the United States, Australia, Canada, England, the Netherlands, Scotland, Sweden, and Wales.

Selection Criteria for Instruments

As noted previously, the emphasis of this review was on standardized scales and interviews that could differentiate between various forms of psychopathology or behavior disorders in persons with mental retardation. The computer search, and more specifically the key word *diagnosis*, produced a very large number of articles that were deemed not to be relevant to this review. These included numerous research papers concerned with identification of various physiological, genetic, metabolic, or other pathological disorders, such as Rett syndrome, phenylketonuria, and so forth. Such publications were excluded from the present review. Also excluded were articles and instruments that attempted to formulate subgroups on the basis of IQ test profiles or neuropsychological profiles. Vocational adaptation and readiness scales were excluded unless specifically relevant to the dual diagnosis question. Finally, scales that were designed to screen for a single disorder, such as the several autism scales, were not included in this review. These criteria were somewhat arbitrary, but it was necessary to put boundaries on the survey so that its major objectives could be achieved.

Another criterion that was applied was that a given instrument needed to be either developed or tested with one or more samples of mentally retarded persons in order to be considered. This, of course, excluded a lot of instruments that were developed for diagnostic purposes in the normal IQ population but which *might* have relevance to persons with mental retardation.

The search resulted in approximately 40 relevant instruments being located. Depending upon the nature of the instrument and its level of development, it was assigned to one of three sections in this review. Part I of the review includes the more established instruments, most of which have been published. These tools were described in detail and thoroughly critiqued. Part II includes relatively new and/or unpublished instruments. The summaries in this section are much shorter, and critiques are often confined to brief statements about the availability or not of various psychometric indices. It was felt that a thorough psychometric critique of these instruments would be more destructive than helpful, as many of these are of recent origin and their developers usually have not had the opportunity to conduct all of the necessary field tests to assess their psychometric properties. Finally, Part III was added so that instruments that were relevant, but peripheral to the assessment of behavior disorders, could be included. This section contains only very brief descriptions of the instruments concerned and no appraisal of their psychometric characteristics.

Instruments Not Included

As noted, several prominent behavior assessment instruments were not reviewed for reasons stated previously. For the interested reader, some of these are listed here. Generally speaking, these instruments are organized by the age group for which they were designed and by type of instrument.

Preschool rating instruments. There are remarkably few of these currently available. The better preschool rating scales include the Problem Checklist (Kohn & Rosman, 1972a, 1972b, and 1973) and the Behavioral Screening Questionnaire developed by Richman and Graham (Earls & Richman, 1980; Richman, Stevenson, & Graham, 1975, 1982). Another useful preschool rating tool is the Preschool Behavior Questionnaire (Behar & Stringfield, 1974a, 1974b), which is described later in this review. Most of the remaining preschool rating scales were developed so long ago that their current utility must be questioned.

Temperament scales. Another group of instruments that have been used primarily to assess preschool and young children are the temperament scales. There are several of these tools available, but perhaps the best known are (1) the scale of temperament used in the New York Longitudinal Study (Thomas & Chess, 1977, 1984), (2) the Infant Temperament Questionnaire (Carey & McDevitt, 1978; McDevitt & Carey, 1978), (3) the Dimensions of Temperament Survey (DOTS) (Lerner, Palermo, Spiro, & Nesselroade, 1982), (4) the Temperament Assessment Battery (Martin, 1984; Paget, Nagle, & Martin, 1984), and (5) the EASI-1 (Buss, Plomin, & Willerman, 1973). Gibbs, Reeves, and Cunningham (1987) have assessed the psychometric properties of several of these; Carey (1982) has commented on their validity; and Hertzig and Snow (1988) have provided an excellent overview of temperament scales.

Scales for school-age children. There are numerous scales available for assessing the general pattern of problem behavior in school-age children, but only some of the most popular ones will be mentioned here. Some instruments, such as the Revised Behavior Problem Checklist (Aman, Werry, Fitzpatrick, Lowe, & Waters, 1983; Quay, 1983; Quay & Peterson, 1983) and the Louisville Behavior Checklist (Miller, 1967) were designed for completion by any responsible adult, usually a parent or teacher. Others, designed solely for completion by parents or primary caretakers, include the Child Behavior Checklist (Achenbach, 1978; Achenbach & Edelbrock, 1979, 1983), Conners' Parent Questionnaire (Conners, 1970, 1973, 1985), the Children's Behavior Questionnaire for Parents (Rutter's *Child Scale A*) (Rutter, Graham, & Yule, 1970), and the Personality Inventory for Children (Kline, Maltz, Lachar, Spector, & Fischoff, 1987; Wirt, Lachar, Klinedinst, & Seat, 1977). Additionally, there are some excellent and well-known scales designed primarily for teacher ratings. These include Conners' Teacher Questionnaire (Conners, 1969, 1973, 1982), the Teacher's Report Form (Achenbach & Edelbrock, 1986), the Children's Behavior Questionnaire for Teachers (Rutter's *Child Scale B*) (Rutter, 1967), and the ADD-H: Comprehensive Teacher Rating Scale (ACTeRS) (Ullmann, Sleator, & Sprague, 1984, 1985). Finally, it should be noted that the Devereux Adolescent Behavior Rating Scale (Spivack, Haimes, & Spotts, 1967) and the Devereux Child Behavior Rating Scale (Spivack & Spotts, 1966) have also been very popular child behavior rating tools, and these are discussed in detail later in the review.

Structured psychiatric interviews. There is also a variety of interviews which attempt to elicit DSM-III, DSM-III-R, or ICD-9 psychiatric symptomatology, where appropriate. These include highly structured interviews, such as the Diagnostic Interview for Children and Adolescents (DICA) (Herjanic & Reich, 1982) and the Diagnostic Interview Schedule (DISC) (Costello, Edelbrock, Dulcan, Kalas, & Klaric, 1984), and semistructured interviews such as the Child Assessment Schedule (Hodges, 1985). In all three instances, there are parallel versions that are worded appropriately both for the parents and the child being rated.

Autism assessment scales. Because of the substantial overlap between childhood autism and mental retardation, some of the better-known instruments for assessing autism are mentioned here. These include diagnostic rating scales such as the Childhood Autism Rating Scale (CARS) (Schopler, Reichler, DeVellis, & Daly, 1980; Schopler, Reichler, & Renner, 1986), the Autism Screening Instrument for Educational Planning (Krug, Arick, & Almond, 1980a, 1980b), and the Diagnostic Checklist for Behavior Disturbed Children (Rimland, 1964, 1968). There are also direct observation systems for assessing the presence or absence of autism, such as the Behavior Observation Scale (BOS) (Freeman et al., 1979; Freeman & Ritvo, 1980; Freeman et al., 1981) and the Behavior Rating Instrument for Autistic and Atypical Children (Ruttenberg, Dratman, Fraknoi, & Wenar, 1966; Ruttenberg, Kalish, Wenar, & Wolf, 1977). Several of the more frequently used methods for assessing autism have been critically assessed in reviews by Morgan (1988) and Parks (1983).

Other Reviews Relevant to the Assessment of Psychopathology

General clinical populations. There are several other reviews that may be of interest to the present readership. Among the better reviews of assessment approaches that are not confined to developmentally disabled populations are those by the following: Achenbach and Edelbrock (1978); Boyle and Jones (1985); Corcoran and Fischer (1987); Dreger (1982); Hammill, Brown, & Bryant (1989); Kestenbaum and Williams (1988); Orvaschel, Sholomskas, and Weissman (1980); Quay (1986); Special Feature on Rating Scales (1985); Sattler (1988); Taylor (1984); and Werry (1978). The discussions by Kestenbaum and Williams, Orvaschel et al., Special Feature on Rating Scales, and Sattler are particularly recommended.

Mentally retarded populations. There are far fewer discussions and critiques of assessment in mental retardation, especially if the focus is narrowed to maladaptive behavior. Some useful discussions include those by Aman and White (1986); Dickens and Stallard (1987); Hogg and Raynes (1987); Mayeda and Lindberg (1980); Meyers, Nihira, and Zetlin (1979); and Walls, Werner, Bacon, and Zane (1977). The reviews by Hogg and Raynes, Mayeda and Lindberg, and Walls et al. are strongly recommended.

Evaluation Criteria

In order to assess the various instruments in a uniform fashion, a standard set of evaluation criteria was adopted. The criteria that were applied to all instruments surveyed in Part I included assessments of the following aspects: (1) Standardization samples employed, (2) Internal consistency, (3) Item-subscale (item-total) correlations, (4) Test-retest reliability, (5) Interrater reliability, (6) Factorial or taxonomic validity, (7) Criterion group validity, and (8) Congruent validity. Most of these are self-explanatory, but a few require further discussion. The standardization samples employed for developing a given tool were noted so that future users of a given instrument will have knowledge of its appropriate application. In general, the writer recommends that instruments not be employed for populations other than those for which they were developed or, if they are so employed, that appropriate caution be exercised in their interpretation. The term *factorial and taxonomy based validity* was used to identify any overarching system used to structure components of the instrument. *Factor validity* is reasonably straightforward and is used here to refer to instruments empirically derived in part or wholly by factor analysis. *Taxonomic validity* was used to refer to a structure for abnormal behavior that usually was extrapolated from one of the widely adopted diagnostic systems, such as those described in the DSM-III-R or the ICD-9. Some of the inherent risks in using diagnostic schemes developed for the population of normal IQ persons will be discussed in a subsequent section.

Criterion group validity was used to refer to comparisons of subjects presumed to have different levels of abnormal behavior. This term frequently was applied rather liberally. For example, comparisons of medicated versus nonmedicated subjects were tabulated and discussed as instances of criterion group validity. Some readers may disagree with the inclusion of some of these comparisons as representative of criterion group validity, but it was felt that it would be better to err on the side of overinclusion.

In addition to the above criteria, if instrument developers made explicit, systematic attempts to address other psychometric issues, these were summarized in narrative form for that instrument. For example, a few authors conducted systematic evaluations of the item content of their instruments by having individual items scrutinized and rated by professionals who had substantial experience in working with mentally retarded persons. These instances were uncommon, but they were pointed out when such instruments were reviewed.

Acceptable Ranges

Many of the statistics cited in this review are correlation coefficients of various types. Of the several measures of internal consistency, such as coefficient alpha and Spearman-Brown coefficients, some authors have indicated that a level of .70 may be satisfactory (e.g., Reiss, 1988). Others have set the lower limit of acceptability at .80 (e.g., Bean & Roszkowski, 1982). In the present review, .70 was adopted as the minimal level for acceptable internal consistency. Levels of .80 and .90 were used to indicate good and excellent levels of internal consistency, respectively.

A host of correlation coefficients, usually Pearson coefficients, are reported in relation to test-retest and interrater reliability. In judging these, it is also helpful to have some qualitative guidelines. A set of commonly adopted reliability levels has been offered by Cicchetti and Sparrow (1981) (following similar suggestions by Fleiss, 1981, and Landis & Koch, 1977). The reliability ranges recommended by Cicchetti and Sparrow are as follows:

Level of Reliability Coefficient	*Clinical Significance*
Less than .40	Poor
.40 to .59	Fair
.60 to .74	Good
.75 to 1.00	Excellent

Of course, these characterizations are somewhat arbitrary, and the evaluation of a given statistic must be tempered by a knowledge of a variety of experimental factors. To help in appreciating the comparisons that are to be presented later in this review, it may be useful to apply these ranges to the rating scale literature involving children of normal IQ. Rating scales have a long tradition of use in clinical research with children of normal IQ, and they often have provided the sole or major means for assignment of children to different clinical groups.

Recently, Achenbach, McConaughy, and Howell (1987) conducted a meta-analysis of the degree of consistency of behavior ratings between different types of informants (parents, teachers, mental health workers, observers, peers, and the subjects themselves) who were involved in interrater reliability studies. Achenbach et al. located 119 relevant studies encompassing 269 samples of children. Studies were excluded if subjects had autism or low IQs (below 50). Achenbach et al. classified the studies in terms of whether or not similar informants (e.g., teacher-teacher pairs), different types of informants (e.g., parent-teacher, teacher-self pairs), or the children themselves conducted the ratings. The data summarized by Achenbach et al. have been reconstructed using the criteria suggested by Cicchetti and Sparrow (1981) and appear in Table 1. It is interesting to note that the modal reliability levels for similar types of informants fall into the cells corresponding to *fair* and *good* reliability levels. In the case of different types of informants, the modal reliability level falls in the cell corresponding to *poor* reliability.

We have conducted this exercise because it provides a frame of reference with which to measure pertinent work in the mental retardation field. Even in the clinical child field, where rating instruments have a long and established role, interrater reliability levels often do not exceed the range of .60 to .74. Furthermore, Achenbach et al. (1987) point out that *low correlations between informants do not necessarily reflect unreliability*. There is also the possibility that different informants contribute validly different information; that is, the children may behave genuinely differently in various settings and in interaction with different informants.

Review Format

In the reviews that appear in Part I, a uniform format was adopted for reporting purposes. The *Point-form Synopsis* was intended to provide an abbreviated summary, so that readers can rapidly scan the features of a given instrument to decide whether or not they wish to read the more detailed summaries. The synopsis also provides certain practical information, such as an instrument's cost and source, should the reader wish to obtain copies. The *Description* sections attempt to relate the history, structure, scoring methods, appropriate users, appropriate subjects, and so forth of each instrument. If an instrument has unique features or conveniences built into its make-up, this was summarized in an *Additional Features* section. Finally, the *Critique* was an attempt to judge each instrument on the evaluation criteria presented above. The critique should be read in conjunction with Table 2 and the summary table appearing in Appendix B. Readers should note that all correlations presented in the summary table (Appendix B) are Pearson product moment correlations unless specifically reported otherwise. Also, readers should note that all citations appearing in Appendix B are referenced in full in their respective sections within Part I.

Some Caveats

When reading the reviews that follow, readers are asked to keep some caveats in mind. First, instruments for which seemingly mediocre psychometric data have been presented may well be preferred over more glamorous-appearing instruments without such data. At least, if such data are available, the professional employing the given tool can be forewarned and make appropriate allowances. Second, the differences between scales in part may reflect varying degrees of candor between different investigators. For example, some workers may be reluctant to report mediocre results, preferring to "improve" their experimental procedures until results more in line with their expectations are obtained. Third, it is apropos to point out that there is no such thing as *the* reliability or validity of a given instrument. The best we can do is to obtain a sample value that, it is to be hoped, is reflective of typical values that can be expected on average with that instrument. Our own studies, which have typically produced a wide range of reliability levels that differ both across raters and subscales, help to highlight this problem (Aman, Singh, Stewart, & Field, 1985; Aman, Singh, & Turbott, 1987). Thus, a simple comparison of statistics across studies may not tell the whole story.

The instruments encompassed within this review differed greatly in terms of their breadth of application. For example, some were designed as simple screening devices for any sort of significant behavior problem, whereas others were much more refined and were developed to render a specific diagnosis. It is important to note that the standards applied for these two types of tools necessarily must differ greatly in terms of their stringency. The developers of a screening instrument may need only to establish that the instrument separates individuals with and without major behavioral problems or psychiatric disorders. On the other hand, developers of diagnosis-specific instruments must attempt to establish the validity of all component dimensions while at the same time showing that the several dimensions do not tend to measure the same thing. Clearly, the development of this type of instrument, while providing adequate evidence of its psychometric integrity, is a much greater challenge than the production of a screening tool. For this reason, more specific diagnostic tools may be faulted more readily in a psychometric review such as this. However, when this does occur, readers should be aware that it eventuates in part because of the higher level of precision aspired to by the tool's makers.

Another point that needs to be made is that several of the instruments reviewed here were never claimed by their developers to be diagnostic instruments for psychopathology. The adaptive behavior scales are a good example, as many of them have maladaptive behavior sections. However, the assessment of inappropriate behaviors was not the major reason for their construction. Nevertheless, these instruments were included in the report in the interests of obtaining coverage as comprehensive as possible.

Finally, in an exercise as extensive as this, it is almost inevitable that some factual errors may have occurred or that some clerical errors may have crept in over the several drafts. If readers or authors of the tests that have been reviewed note any factual errors, the writer asks that these be brought to his attention. Likewise, although an earnest effort was made to locate and include all relevant instruments, it is likely that some appropriate materials were missed. Again, the writer asks that any such omissions be brought to his attention. From time to time, we hope to update this review, and feedback of this type will be very helpful in ensuring accuracy, comprehensiveness, and balance in future endeavors.

The Nature of Psychopathology in Mental Retardation

One final issue must be addressed before launching into the review of available scales, and that involves the very nature of psychopathology when it occurs among persons with mental retardation. These days, it is common to read that the full range of psychopathology can be found in mentally retarded persons. It is also common to see diagnostic surveys in which an established diagnostic system (such as the DSM-III-R) is used, apparently successfully, to classify the disorders presented by disturbed individuals with mental retardation. However, this in no way validates these diagnostic systems as the correct taxonomic system for classifying behavior disorders in mentally retarded persons.

The writer has assumed a position that perhaps may prove to be both unpopular and controversial; namely, that *the application of established diagnostic schemes is increasingly suspect as the severity of the patient's mental retardation increases*. However, it only makes sense that the stresses affecting a person, his or her appraisal of those stressors, and the ultimate *expression* of psychopathology may take on very different forms in individuals having substantial intellectual handicaps. Indeed, the DSM-III-R deals explicitly with just this type of problem when discussing the use of its diagnostic guidelines with different cultures:

> When the DSM-III-R classification and diagnostic criteria are used to evaluate a person from an ethnic or cultural group different from that of the clinician's,...caution should be exercised in the application of DSM-III-R diagnostic criteria to assure that their use is culturally valid. It is important that the clinician not employ DSM-III-R in a mechanical fashion, insensitive to differences in language, values, behavioral norms, and idiomatic expression of distress. (APA, 1987, p. xxvi).

Of course, the same can be said of any psychiatric taxonomic system developed on a population overwhelmingly made up of normal IQ people. The point I wish to make here is that the presence of a substantial intellectual handicap may be functionally equivalent to and probably even more profound than the cultural barriers alluded to in the DSM-III-R caveat. This should be recognized, notwithstanding the desirability and the enormous gains made in striving for normalization in recent years.

A number of other workers have commented on the enormous problems in applying established diagnostic systems to persons with mental retardation. Reid (1983), for example, has discussed a number of impediments to achieving accurate diagnoses in disturbed mentally retarded people. For example, the lack of speech or the presence of concrete speech may make it very difficult to determine the presence of certain symptoms, such as delusions, hallucinations, extreme affect, and so forth. Furthermore, the presence of certain behaviors (e.g., echolalia, stereotypy), which ordinarily would be deemed as abnormal in people of normal IQ, may be developmentally appropriate in persons of low mental age (Reid, 1983). All of these considerations seem to challenge the routine application of traditional taxonomic psychiatric systems across the spectrum of mental retardation. On the other hand, the use of such systems would seem to be appropriate among persons with borderline intelligence, mild mental retardation, and (possibly) moderate mental retardation. For all of these reasons, the use of established systems has been reviewed in this report as probably appropriate when confined to higher functional levels. However, in the reviews to follow, *it has been judged as potentially invalid and in need of supporting evidence* when this approach has been applied to a broader spectrum of developmental handicap.

A final point concerns the establishment of validity in an area such as behavioral/psychiatric diagnosis in mental retardation where *no gold standard already exists*. As Achenbach and Edelbrock (1978) have noted, one usually develops a new instrument because of dissatisfaction with the preexisiting array of tools. This creates special problems when it comes to validating new instruments due to a lack of suitable comparison methods. Achenbach and Edelbrock were referring to the clinical child literature when they raised this issue, but the dilemma would seem to be even more complex in the mental retardation field.

References

Achenbach, T. M. (1978). The child behavior profile: I. Boys aged 6-11. *Journal of Consulting and Clinical Psychology, 46*, 478-488.

Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin, 85*, 1275-1301.

Achenbach, T. M., & Edelbrock, C. S. (1979). The child behavior profile: II. Boys aged 12-16 and girls aged 6-11 and 12-16. *Journal of Consulting and Clinical Psychology, 47*, 223-233.

Achenbach, T. M., & Edelbrock, C. S. (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile.* Burlington, VT: Department of Psychiatry, University of Vermont.

Achenbach, T. M., & Edelbrock, C. S. (1986). *Manual for the Teacher's Report Form and teacher version of the Child Behavior Profile.* Burlington, VT: Department of Psychiatry, University of Vermont.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross informant correlation for situational specificity. *Psychological Bulletin, 101*, 213-232.

Aman, M. G., Werry, J. S., Fitzpatrick, J., Lowe, M., & Waters, J. (1983). Factor structure and norms for the Revised Behavior Problem Checklist in New Zealand children. *Australian and New Zealand Journal of Psychiatry, 17*, 354-360.

Aman, M. G., Singh, N. N., Stewart, A. W., & Field, C. J. (1985). Psychometric characteristics of the Aberrant Behavior Checklist. *American Journal of Mental Deficiency, 89*, 492-502.

Aman, M. G., & White, A. J. (1986). Measures of drug change in mental retardation. In K. D. Gadow (Ed.), *Advances in learning and behavioral disabilities* (Vol. 5, pp. 157-202). Greenwich, CT: JAI Press.

Aman, M. G., Singh, N. N., & Turbott, S. H. (1987). Reliability of the Aberrant Behavior Checklist and the effect of variations in instructions. *American Journal of Mental Deficiency, 92*, 237-240.

American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., revised). Washington, DC: Author.

Bean, A. B., & Roszkowski, M. J. (1982). Item-domain relationships in the Adaptive Behavior Scale (ABS). *Applied Research in Mental Retardation, 3*, 359-367.

Behar, L., & Stringfield, S. (1974a). A behavior rating scale for the preschool child. *Developmental Psychology, 10*, 601-610.

Behar, L., & Stringfield, S. (1974b). *Manual for the Preschool Behavior Questionnaire.* Unpublished manuscript, Children's Mental Health Services, Division of Mental Health Services, Durham, NC.

Boyle, M. H., & Jones, S. C. (1985). Selecting measures of emotional and behavioral disorders of childhood for use in general populations. *Journal of Child Psychology and Psychiatry, 26*, 137-159.

Buss, A. H., Plomin, R., & Willerman, L. (1973). The inheritance of temperament. *Journal of Personality, 41*, 513-524.

Carey, W. B., & McDevitt, S. C. (1978). Revision of the Infant Temperament Questionnaire. *Pediatrics, 61*, 735-739.

Carey, W. B. (1982). Validity of parental assessments of development and behavior. *American Journal of Diseases of Children, 136*, 97-99.

Cicchetti, D. V., & Sparrow, S. S. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*, 127-137.

Conners, C. K. (1969). A teacher rating scale for use in drug studies with children. *American Journal of Psychiatry, 126*, 152-156.

Conners, C. K. (1970). Symptom patterns in hyperkinetic, neurotic, and normal children. *Child Development, 41*, 667-682.

Conners, C. K. (1973). Rating scales for use in drug studies with children. *Psychopharmacology Bulletin* (Special Issue, Pharmacotherapy with Children), 24-84.

Conners, C. K. (1982). Parent and teacher rating forms for the assessment of hyperkinesis in children. In P. A. Keller & L. G. Ritt (Eds.), *Innovations in clinical practice: A source book* (Vol. 1, pp. 257-264). Sarasota, FL: Professional Resource Exchange, Inc.

Conners, C. K. (1985). Parent Symptom Questionnaire. *Psychopharmacology Bulletin, 21*, 816-822.

Corcoran, K., & Fischer, J. (1987). *Measures for clinical practice: A sourcebook.* New York: Free Press.

Costello, A. J., Edelbrock, C., Dulcan, M. K., Kalas, R., & Klaric, S. H. (1984). *Development and testing of the NIMH diagnostic interview schedule for children in a clinic population.* (Contract #RFP-DB-81-0027). Center for Epidemiologic Studies, NIMH, Rockville, MD.

Dickens, P., & Stallard, A. (1987). *Assessing mentally handicapped people: A guide for care staff.* Philadelphia: NFER-Nelson Publishing Company.

Dreger, R. M. (1982). The classification of children and their emotional problems: An overview--II. *Clinical Psychology Review, 2*, 349-385.

Earls, F., & Richman, N. (1980). Behavior problems in pre-school children of West Indian-born parents: A re-examination of family and social factors. *Journal of Child Psychology and Psychiatry, 21*, 107-117.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.

Freeman, B. J., Guthrie, D., Ritvo, E., Schroth, P., Glass, R., & Frankel, F. (1979). Behavior Observation Scale: Preliminary analysis of the similarities and differences between autistic and mentally retarded children. *Psychological Reports, 44*, 519-524.

Freeman, B. R., & Ritvo, E. R. (1980). Behavior Observation Scale for autism (BOS). *International Journal of Rehabilitation Research, 3*, 254-258.

Freeman, B. R., Ritvo, E. R., Schroth, P. C., Tonick, I., Guthrie, D., & Wake, L. (1981). Behavioral characteristics of high- and low-IQ autistic children. *American Journal of Psychiatry, 138*, 25-29.

Gibbs, M. V., Reeves, D., & Cunningham, C. C. (1987). The application of temperament questionnaires to a British sample: Issues of reliability and validity. *Journal of Child Psychology and Psychiatry, 28*, 61-77.

Hammill, D. D., Brown, L., & Bryant, B. R. (1989). *A consumer's guide to tests in print.* Austin, TX: Pro-ed.

Herjanic, B., & Reich, W. (1982). Development of a structured psychiatric interview for children: Agreement between child and parent on individual symptoms. *Journal of Abnormal Child Psychology, 10*, 307-324.

Hertzig, M. E., & Snow, M. E. (1988). The assessment of temperament. In C. J. Kestenbaum & D. T. Williams (Eds.), *Handbook of clinical assessment of children and adolescents* (Vol. I, pp. 133-153). New York: New York University Press.

Hodges, K. (1985). *Manual for the Child Assessment Schedule (CAS).* Unpublished manuscript, University of Missouri, Columbia.

Hogg, J., & Raynes, N. N. (Eds.) (1987). *Assessment in mental handicap: A guide to assessment practices, tests, and checklists* . London: Croom Helm.

Kestenbaum, C. J., & Williams, D. T. (Eds.) (1988). *Handbook of clinical assessment of children and adolescents* (Vol. I). New York: New York University Press.

Kline, R. B., Maltz, A., Lachar, D., Spector, S., & Fischoff, J. (1987). Differentiation of infantile autistic, child-onset pervasive developmental disorder, and mentally retarded children, with the Personality Inventory for Children. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 839-843.

Kohn, M., & Rosman, B. L. (1972a). Relationship of preschool social-emotional functioning to later intellectual achievement. *Developmental Psychology, 6*, 445-452.

Kohn, M., & Rosman, B. L. (1972b). A social competence scale and symptom checklist for the preschool child: Factor dimensions, their cross instrument generality, and longitudinal persistence. *Developmental Psychology, 6*, 430-444.

Kohn, M., & Rosman, B. L. (1973). A two-factor model of emotional disturbance in the young child: Validity and screening efficiency. *Journal of Child Psychology and Psychiatry, 14*, 31-56.

Krug, D. A., Arick, J. R., & Almond, P. J. (1980a). *Autism Screening Instrument for Educational Planning.* Portland, OR: ASIEP Educational.

Krug, D. A., Arick, J., & Almond, P. (1980b). Behavior checklist for identifying severely handicapped individuals with high levels of autistic behavior. *Journal of Child Psychology and Psychiatry, 21*, 221-229.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Lerner, R. M., Palermo, M., Spiro, A. III, & Nesselroade, J. R. (1982). Assessing the dimensions of temperamental individuality across the life span: The Dimensions of Temperament Survey (DOTS). *Child Development, 53*, 149-159.

Martin, R. P. (1984). *The Temperament Assessment Battery interim manual.* Athens, GA: Developmental Metrics.

Mayeda, T., & Lindberg, G. (1980). *Performance measures of skill and adaptive competencies: Annotations of measures of potential use in developmental disabilities programs.* Los Angeles: University of California Neuropsychiatric Institute Research Group at Lanterman State Hospital and Developmental Center.

McDevitt, S. C., & Carey, W. (1978). The measurement of temperament in 3-7 year old children. *Journal of Child Psychology and Psychiatry, 19,* 245-253.

Meyers, C. E., Nihira, K., & Zetlin, A. (1979). The measurement of adaptive behavior. In N. R. Ellis (Eds.), *Handbook of mental deficiency* (2nd ed., pp. 431-481). New York: Lawrence Erlbaum Associates.

Miller, L. (1967). Louisville Behavior Checklist for males 6-12 years of age. *Psychological Reports, 21,* 885-896.

Morgan, S. (1988). Diagnostic assessment of autism: A review of objective scales. *Journal of Psychoeducational Assessment, 6,* 139-151.

Orvascel, H., Sholomskas, D., & Weissman, M. M. (1980). *The assessment of psychopathology and behavior problems in children: A review of scales suitable for epidemiological and clinical research (1967-1979).* DHHS Publication No. (ADM)80-1037, Superintendent of Documents, Washington, DC: U.S. Government Printing Office.

Paget, K. D., Nagle, R. J., & Martin, R. P. (1984). Interrelationships between temperament characteristics and first-grade teacher-student interactions. *Journal of Abnormal Child Psychology, 12,* 547-560.

Parks, S. (1983). The assessment of autistic children: A selective review of available instruments. *Journal of Autism and Developmental Disorders, 13,* 255-267.

Quay, H. C. (1983). A dimensional approach to children's behavior disorders: The Revised Behavior Problem Checklist. *School Psychology Review, 12,* 244-249.

Quay, H. C. (1986). Classification. In H. C. Quay & J. S. Werry (Eds.), *Psychopathological disorders of childhood* (3rd ed., pp. 1-34). New York: John Wiley & Sons.

Quay, H. C., & Peterson, D. R. (1983). *Manual for the Revised Behavior Problem Checklist.* Unpublished manual available from H. C. Quay, Box 240074, University of Miami, Coral Gables, FL 33124.

Reid, A. H. (1983). Psychiatry of mental handicap: A review. *Journal of the Royal Society of Medicine, 76,* 587-592.

Reiss, S. (1988). *Test manual for the Reiss Screen for Maladaptive Behavior.* Oakland Park, IL: International Diagnostic Systems.

Reiss, S. (Ed.), (1989). Mini-series: Mental health aspects of mental retardation: Results of NIMH-UIC Conference. [Special issue]. *Research in Developmental Disabilities, 10*(3).

Richman, N., Stevenson, J. E., & Graham, P. J. (1975). Prevalence of behaviour problems in 3-year-old children: An epidemiological study in a London borough. *Journal of Child Psychology and Psychiatry, 16*, 277-287.

Richman, N., Stevenson, J., & Graham, P. J. (1982). *Pre-school to school: A behavioural study.* London: Academic Press.

Rimland, B. (1964). *Infantile autism: The syndrome and its implications for a neural theory of behavior.* New York: Appleton-Century-Crofts.

Rimland, B. (1968). On the objective diagnosis of infantile autism. *Acta Paedopsychiatrica, 35*, 146-161.

Ruttenberg, B. A., Dratman, M. L., Fraknoi, J., Wenar, C. (1966). An instrument for evaluating autistic children. *Journal of the American Academy of Child Psychiatry, 5*, 453-478.

Ruttenberg, B. A., Kalish, B. I., Wenar, C., & Wolf, E. G. (1977). *Behavior rating instrument for autistic and other atypical children* (Rev. ed.). Philadelphia Developmental Center for Autistic Children.

Rutter, M. (1967). Children's behaviour questionnaire for completion by teachers. Preliminary findings. *Journal of Child Psychology and Psychiatry, 8*, 1-11.

Rutter, M., Graham, P., & Yule, W. (1970). *A neuropsychiatric study in childhood.* London: Spastics International Publications.

Sattler, J. (1988). *Assessment of children* (3rd ed.). San Diego: Author.

Schopler, E., Reichler, R. J., DeVellis, R. F., & Daly, K. (1980). Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of Autism and Developmental Disorders, 10*, 91-103.

Schopler, E., Reichler, R. J., & Renner, R. (1986). *The Childhood Autism Rating Scale (CARS) for diagnostic screening and classification of autism.* New York: Irvington Publishers.

Special feature on rating scales and assessment instruments for use in pediatric psychopharmacology research. (1985). *Psychopharmacology Bulletin, 21*, 713-1125.

Special feature on treatment research with the mentally retarded, mentally ill. (1986). *Psychopharmacology Bulletin, 22*, 1045-1087.

Spivack, G., Haimes, P. E., & Spotts, J. (1967). *Devereux Adolescent Behavior Rating Scale manual.* Devon, PA: The Devereux Foundation.

Spivack, G., & Spotts, J. (1966). *Devereux Child Behavior Rating Scale manual.* Devon, PA: The Devereux Foundation Press.

Taylor, R. L. (1984). *Assessment of exceptional students: Educational and psychological procedures.* Englewood Cliffs, NJ: Prentice-Hall.

Thomas, A. & Chess, S. (1977). *Temperament and development.* New York: Brunner/Mazel.

Thomas, A., & Chess, S. (1984). Genesis and evolution of behavioral disorders: From infancy to early adult life. *American Journal of Psychiatry, 141*, 1-9.

Ullmann, R. K., Sleator, E. K., & Sprague, R. L. (1984). A new rating scale for diagnosis and monitoring of ADD children. *Psychopharmacology Bulletin, 20*, 160-164.

Ullmann, R. K., Sleator, E. K., & Sprague, R. L. (1985). Introduction to the use of ACTeRS. *Psychopharmacology Bulletin, 21*, 915-916.

Walls, R. T., Werner, T. J., Bacon, A., & Zane, T. Behavior checklists. In J. D. Cone & R. P. Hawkins (Eds.). *Behavioral assessment: New directions in clinical psychology* (pp. 77-146). New York: Brunner/Mazel, 1977.

Werry, J. S. (1978). Measures in pediatric psychopharmacology. In J. S. Werry (Ed.), *Pediatric psychopharmacology. The use of behavior modifying drugs in children* (pp. 29-78). New York: Brunner/Mazel.

Wirt, R. D., Lachar, D., Klinedinst, J. K., & Seat, P. D. (1977). *Multidimensional description of child personality: A manual for the Personality Inventory for Children.* Los Angeles: Western Psychological Services.

Part I

Detailed Summaries and Critiques:
Published and More Established Instruments

AAMD Adaptive Behavior Scale: Residential and Community Edition
(Part II)

K. Nihira, R. Foster, M. Shellhaas, & H. Leland, 1975

Point-form Synopsis

Stated Purpose: To provide objective descriptions and evaluations of adaptive behavior, defined as the individual's effectiveness in coping with the natural and social demands of his or her environment.

Age Range: Early childhood through late adulthood.

Level of Mental Retardation Covered: Mild through profound.

Raters/Diagnosers: Both professionals and nonprofessionals having substantial experience with the individual being rated.

Time Required to Complete (Part II): Estimated by reviewer at 15 to 30 minutes.

Disorders/Dimensions Identified (Part II): Fourteen domains are scored as follows: (1) Violent & Destructive Behavior, (2) Antisocial Behavior, (3) Rebellious Behavior, (4) Untrustworthy Behavior, (5) Withdrawal, (6) Stereotyped Behavior & Odd Mannerisms, (7) Inappropriate Interpersonal Manners, (8) Unacceptable Vocal Habits, (9) Unacceptable or Eccentric Habits, (10) Self-Abusive Behavior, (11) Hyperactive Tendencies, (12) Sexually Aberrant Behavior, (13) Psychological Disturbances, and (14) Use of Medications.

Date of Manual Publication: 1975.

Cost: Manual, $10.00. Package of 10 test booklets, $20.00; 100 booklets, $120.

Source: Pro Ed Inc., 5341 Industrial Oaks Boulevard, Austin, TX 78735. Telephone (512) 451-3246; FAX (512) 451-8542.

Limitations/Exclusions: Published norms available only for institutionalized populations.

Description

The AAMD Adaptive Behavior Scale: Residential and Community Edition (hereinafter simply called the Adaptive Behavior Scale [ABS]) is an informant instrument designed to assess the adaptive behavior of mentally retarded, emotionally maladjusted, and developmentally disabled individuals. Adaptive behavior is defined as the effectiveness of an individual in coping with the natural and social demands of his or her environment. The ABS is divided into two major parts. Part I is organized developmentally and is intended to evaluate the individual's skills and habits in 10 behavioral domains regarded as being important in achieving personal independence in daily living. The Part I domains and, where relevant, the number of subdomains are as follows: (1) Independent Functioning (8 subdomains), (2) Physical Development (2 subdomains), (3) Economic Activity (2 subdomains), (4) Language Development (3 subdomains), (5) Numbers and Time, (6) Domestic Activity (3 subdomains), (7) Vocational Activity, (8) Self-Direction (3 subdomains), (9) Responsibility, and (10) Socialization. All in all, Part I comprises 66 questions, which are further broken down into a total of 351 component statements or choices. Higher scores signify a higher level of adaptive functioning on all of the Part I domains.

Part II is broken down into 14 domains, and many of these are further divided into subdomains as follows: (1) Violent and Destructive Behavior (5 subdomains), (2) Antisocial Behavior (6 subdomains), (3) Rebellious Behavior (6 subdomains), (4) Untrustworthy Behavior (2 subdomains), (5) Withdrawal (3 subdomains), (6) Stereotyped Behavior & Odd Mannerisms (2 subdomains), (7) Inappropriate Interpersonal Manners, (8) Unacceptable Vocal Habits, (9) Unacceptable or Eccentric Habits (4 subdomains), (10) Self-Abusive Behavior, (11) Hyperactive Tendencies, (12) Sexually Aberrant Behavior (4 subdomains), (13) Psychological Disturbances (7 subdomains), and (14) Use of Medication. Each subdomain comprises a group of apparently similar items. For example, item number 22, which relates to shyness under the Withdrawal domain, has the following components: (a) Is timid and shy in social situations; (b) Hides face in group situations; (c) Does not mix well with others; (d) Prefers to be alone; and (e) Other (specify). The individual items are left unscored if they do not apply to the subject, and they are scored as "1" if they occur occasionally or "2" if they occur frequently. Higher scores in Part II signify more numerous behavior problems on the given domain.

The manual indicates that the ABS can be completed both by professionals and nonprofessionals. The professions mentioned include psychologists, social workers, speech and hearing personnel, and so forth. With appropriate supervision, any responsible person can complete the scale, including institutional aides and nurses, parents, outreach workers, teachers, and workshop supervisors. A variety of uses for the ABS are suggested in the manual, as follows: (1) to identify areas of deficiency needing to be addressed, (2) to provide a basis for comparison over time, (3) to assess the same individual in different settings, (4) to assess differences in rater-subject relationships, (5) to enhance the exchange of information by providing a standardized reporting system, and (6) to facilitate administrative decision making.

Items for Part I apparently resulted in part from a review of existing behavior rating scales and *a priori* assignment of these to their respective domains. Items for Part II were derived from a critical incident study in which psychiatric aides, special education teachers, and attendants in day care centers reported behaviors of mentally retarded subjects that were considered unacceptable. These behaviors subsequently were classified into categories by judges. When disagreements occurred, a given incident (item) was reclassified until agreement was attained (Nihira, 1973). Thus, allocation of items to the various domains appears to have been achieved by consensus.

Additional Features

The 1975 manual makes reference to a Fortran computer program and key punch format for machine scoring and organization of the data using the 1969 edition of the ABS.

Critique

Only Part II of the ABS will be reviewed here, because of the concern of this report with behavior disorders. Furthermore, data will not be reported for the Use of Medications domain, as this is not a description of a behavioral symptom or pattern. The psychometric characteristics of the ABS, Part II, are summarized in Table 2 and Appendix B. Spreat (1982a) has also reviewed this scale in detail. The ABS was relatively well standardized, with data available on over 4,000 subjects, aged 3 to 69 years. However, standardization data are only provided for institutionalized individuals. Given the popularity of this instrument, there are surprisingly few psychometric data. The writer was able to locate only one study of the instrument's internal consistency, this being a study by Bean and Roszkowski (1982). Alpha coefficients ranged from a low of .64 to a high of .92, with a mean of .78. This can be regarded as acceptable overall, although not high (Nunnally, 1967). Bean and Roszkowski also examined item-total correlations for the ABS, Part II. Overall, 62% of items were judged as having good item-total consistency, but the remainder were regarded as unsatisfactory (i.e., correlating less than .30 with their own domain or correlating higher with other domains). Rather surprisingly, the manual reports no test-retest reliability data for the ABS. A report by Isett and Spreat (1979) is the only one that could be located which addressed this issue. Two-week test-retest reliability (Spearman) coefficients ranged from .60 to .97 across domains (mean=.83), levels which are generally acceptable to very good. Interrater reliability has been addressed in at least four studies, and mean correlations ranged from .49 to .56 (Nihira, Foster, Shellhaas, & Leland, 1975; Isett and Spreat, 1979; Salagaras and Nettelbeck, 1983; Stack, 1984). These may be acceptable, especially for the more reliable domains, but in general they are cause for some concern.

Much still remains to be determined insofar as the validity of the ABS, Part II, is concerned. As noted, items were allocated to domains on an *a priori* basis. Although this may provide some evidence for its content validity, there is little support for the factorial/taxonomic validity of Part II. Several factor analyses have been done (see Nihira et al., 1975), and these generally indicate separate Personal Maladaption and Social Maladaption factors on Part II. However, these analyses were performed on domain (rather than individual item) scores, so they really do not address the question of appropriate assignment of questions to subscales. In the only factor analysis known to the reviewer to analyze at the level of individual questions, there was a poor match-up between empirically derived factors and existing domains (Nihira, 1978). There is a modicum of criterion group validity with the ABS, Part II. Nihira et al. (1975) found that some domains discriminated between subjects placed in different units, but details were sketchy. Spreat (1980) found that a combination of Part I and Part II domains could differentiate beyond chance levels between subjects in different administrative placements. In keeping with the above comments on factorial validity, Spreat found that empirically derived factors were more accurate than preexisting domains in classifying subjects. Salagaras and Nettelbeck (1983) found that subjects from certain criterion groups tended to score better than others. For example, subjects with Down syndrome, those residing in smaller residential settings, and those not taking medication received significantly lower domain scores than their counterparts. There is only a small amount of congruent validity data with the instrument. Clements, DuBois, Bost, and Bryan (1981) found that global ratings of behavior disturbance were correlated, although somewhat weakly, with ABS Part II total scores. This improved somewhat when a correction to weight items according to their severity was employed. Finally, Aman, Singh, Stewart, and Field (1985) observed significant correspondences between several, although not all, ABS domains which had analogous subscales on the Aberrant Behavior Checklist.

There has been considerable debate about the scoring system employed in the ABS, Part II. A number of authors have argued that the frequency format employed does not reflect adequately the marked differences in the severity of the symptoms described (e.g., Clements, Bost, DuBois, & Turpin, 1980; Clements, DuBois, Bost, & Bryan, 1981; McDevitt, McDevitt, & Rosen, 1977; Holmes & Batt, 1980; MacDonald & Barton, 1986; Taylor, Warren, & Slocumb, 1979). There are some data suggesting that a weighted scoring system may be more valid (e.g., Clements et al., 1981), although others have observed essentially no differences between weighted and unweighted formats (e.g., Searls, Isett, & Bowders, 1981; Spreat, 1982b).

In summary, the ABS is a relatively well-standardized instrument when compared with others in the field, although normative values in the 1975 manual are based solely on institutional populations. The internal consistency of Part II items appears to be acceptable, but item-domain correlations suggest that some items may be misclassified. Test-retest reliability is good, but interrater reliability appears to be marginal, especially for the less reliable domains. There is a general lack of evidence concerning the ABS's factorial validity, and the only relevant factor analysis appears to conflict with the placement of many items. There is a small amount of data on the criterion group and congruent validity of the ABS. However, the validity of a number of domains has not been addressed thus far, and much more needs to be done to establish the technical merits of all domains. The ABS is one of the most popular instruments in the mental retardation field, and it is regrettable that more is not known about its psychometric characteristics. It must be concluded that the value of specific domains for identifying subjects with particular types of behavior disorders is largely unknown. However, at the time of this writing, the ABS is under revision, and the new version is expected to be released in 1990 (H. Leland, personal communication, 11 October 1989). It is possible that the new ABS will resolve a number of the questions raised here.

References

Aman, M. G., Singh, N. N ., Stewart, A. W., & Field, C. J. (1985). Psychometric characteristics of the Aberrant Behavior Checklist. *American Journal of Mental Deficiency*, *89*, 492-502.

Bean, A. G., & Roszkowski, M. J. (1982). Item-domain relationships in the Adaptive Behavior Scale (ABS). *Applied Research in Mental Retardation*, *3*, 359-367.

Clements, P. R., Bost, L. W., DuBois, Y. G., & Turpin, W. B. (1980). Adaptive Behavior Scale, Part Two: Relative severity of maladaptive behavior. *American Journal of Mental Deficiency, 84*, 465-469.

Clements, P. R., Bost, L. W., DuBois, Y. G., & Bryan, C. (1981). Adaptive Behavior Scale, Part Two: Predictive efficiency of severity and frequency scores. *American Journal of Mental Deficiency, 85*, 433-434.

Holmes, C. B., & Batt, R. (1980). Is choking others really equivalent to stamping one's feet? An analysis of Adaptive Behavior Scale items. *Psychological Reports, 46*, 1277-1278.

Isett, R. D., & Spreat, S. (1979). Test-retest and interrater reliability of the AAMD Adaptive Behavior Scale. *American Journal of Mental Deficiency, 84*, 93-95.

MacDonald, L., & Barton, L. E. (1986). Measuring severity of behavior: A revision of Part II of the Adaptive Behavior Scale. *American Journal of Mental Deficiency, 90*, 418-424.

McDevitt, S. C., McDevitt, S. C., & Rosen, M. (1977). Adaptive Behavior Scale, Part II: A cautionary note and suggestions for revisions. *American Journal of Mental Deficiency, 85*, 210-212.

Meyers, C. E., Nihira, K., & Zetlin, A. (1979). The measurement of adaptive behavior. In N. R. Ellis (Ed.), *The handbook of mental deficiency* (2nd ed., pp. 431-481). New York: Lawrence Erlbaum Associates.

Nihira, K. (1973). Importance of environmental demands in the measurement of adaptive behavior. In G. Tarjan, R. K. Eyman, & C. E. Myers (Eds.) *Sociobehavioral studies in mental retardation* (pp. 101-116). Monographs of American Association on Mental Deficiency #1.

Nihira, K. (1978, May). *Dimensions of maladaptive behavior in institutionalized mentally retarded persons.* Paper presented at the 4th annual Western Conference on Mental Retardation, "The Profoundly Mentally Retarded," Corpus Christi, TX.

Nihira, K., Foster, R., Shellhaas, M., & Leland, H. (1975). *AAMD Adaptive Behavior Scale Manual* (rev). Washington, D. C.: American Association on Mental Deficiency.

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.

Roszkowski, M. J. (1980). Concurrent validity of the Adaptive Behavior Scale as assessed by the Vineland Social Maturity Scale. *American Journal of Mental Deficiency, 85*, 86-89.

Salagaras, S. & Nettelbeck, T. (1983). Adaptive behavior of mentally retarded adolescents attending school. *American Journal of Mental Deficiency, 88*, 57-68.

Spreat, S. (1980). The Adaptive Behavior Scale: A study of criterion validity. *American Journal of Mental Deficiency, 85*, 61-68.

Spreat, S. (1982a). The AAMD Adaptive Behavior Scale: A psychometric review. *Journal of School Psychology, 20*, 45-56.

Spreat, S. (1982b). An empirical analysis of item weighting on the Adaptive Behavior Scale. *American Journal of Mental Deficiency, 87*, 159-163.

Stack, J. G. (1984). Interrater reliabilities of the Adaptive Behavior Scale with environmental effects controlled. *American Journal of Mental Deficiency, 88*, 396-400.

Taylor, R. L., Warren, S. A., & Slocumb, P. R. (1979). Categorizing behavior in terms of severity: Considerations for Part Two of the Adaptive Behavior Scale. *American Journal of Mental Deficiency, 83*, 411-414.

AAMD Adaptive Behavior Scale:
School Edition
N. Lambert, M. Windmiller, D. Tharinger, & L. Cole, 1981

Point-form Synopsis

Stated Purpose: To aid school personnel in obtaining measures of personal independence and social skills and to reveal areas of functioning for which special educational program planning may be required.

Age Range: Children aged 3 to 16 years, inclusive.

Level of Mental Retardation Covered: This instrument was designed for assessing children in Regular, Educable Mentally Retarded (EMR) and Trainable Mentally Retarded (TMR) classes. The EMR and TMR designations correspond roughly to mild and moderate-to-severe mental retardation, respectively.

Raters/Diagnosers: Any adult who has a good knowledge of the child can serve as an informant.

Time Required to Complete: When first-person assessment is employed, completion of entire instrument is reported to take 15 to 45 minutes. When third-party assessment is employed, rating time will be longer.

Disorders/Dimensions Identified (Part II):

Twelve domains are scored as follows: (1) Aggressiveness, (2) Antisocial vs. Social Behavior, (3) Rebelliousness, (4) Trustworthiness, (5) Withdrawal vs. Involvement, (6) Mannerisms, (7) Interpersonal Manners, (8) Acceptability of Vocal Habits, (9) Acceptability of Habits, (10) Activity Level, (11) Symptomatic Behavior, and (12) Use of Medications.

Date of Manual Publication: 1981.

Cost: ABS:SE starter set (1 administration and instructional planning manual, 1 diagnostic and technical manual, 2 assessment booklets, 2 instructional planning profiles, 2 diagnostic profiles and 2 parents guides), $31.00; 20 assessment booklets, $25.00; 20 instructional planning profiles, $10.00.

Source: Pro-Ed, 8700 Shoal Creek Boulevard, Austin TX, 78758-9965.
Telephone (512) 451-3246; FAX (512) 451-8542.

Limitations/Exclusions: No standardization or norms for parent ratings. Relevance of norms to children with profound mental retardation uncertain.

Description

The AAMD Adaptive Behavior Scale, School Edition (ABS:SE) is based on the Adaptive Behavior Scale, Public School Version, which in turn was derived from the AAMD Adaptive Behavior Scale: Residential and Community Edition (described in the preceding section of this review). The ABS:SE was developed to aid school personnel to assess children's personal independence and social skills and to reveal areas of functioning requiring special program planning (Lambert, Windmiller, Tharinger, & Cole, 1981). The manual contains norms for children aged from 3 through 16 years of age inclusive, and broken down by educational classification: Regular, Educable Mentally Retarded (EMR) and Trainable Mentally Retarded (TMR). However, average domain scores are not available for EMR children below 7 years of age or regular class students over the age of 15 years.

The ABS:SE has two parts comprising a total of 21 behavioral domains. Part I is organized along developmental lines and is intended to assess a person's skills and habits in nine areas, which were taken from its close relative, the ABS: Residential and Community Edition. These domains bear the following names: (1) Independent Functioning, (2) Physical Development, (3) Economic Activity, (4) Language Development, (5) Numbers and Time, (6) Prevocational Activity, (7) Self-Direction, (8) Responsibility, and (9) Socialization. Part II contains 12 domains intended to assess adaptive behavior related to personality and behavior disorders. The Part II domains are labeled as follows: (1) Aggressiveness, (2) Antisocial vs. Social Behavior, (3) Rebelliousness, (4) Trustworthiness, (5) Withdrawal vs. Involvement, (6) Mannerisms, (7) Appropriateness of Interpersonal Manners, (8) Acceptability of Vocal Habits, (9) Acceptability of Habits, (10) Activity Level, (11) Symptomatic Behavior, and (12) Use of Medications. Unlike the ABS: Residential and Community Edition, high scores on

Part II of the ABS:SE are indications of relatively trouble-free behavior. Thus, a child with Part II domain scores *below* the 5th or 10th percentile would be expected to have fairly marked or serious behavior problems on that particular domain.

The ABS:SE comprises 95 items taken from the ABS:Residential and Community Edition. Items not included in the ABS:SE were those judged by teachers, special education experts, pupil personnel, and research staff members as not readily observable in the school setting. This resulted in the deletion of the Domestic Activity domain from Part I and Self-Abusive Behavior and Sexually Aberrant Behavior from Part II. As the ABS:SE is an outgrowth of the earlier ABS, its structure and rationale for assignment of items to domains is basically the same. Items for Part I were taken from a review of existing scales and were assigned on an *a priori* basis to domains. Items for Part II originally were derived from a critical incident study of problematic behaviors filled in by day care staff and teachers, and items were assigned by judges into their respective categories (Nihira, 1973). Unlike the ABS: Residential and Community Edition, the materials for the ABS:SE make it possible to calculate five factor scores that were empirically derived. There have been several factor analytic studies of the ABS:SE, and these generally have produced two or three dimensions on Part I and two dimensions on Part II (Lambert, 1981). These factors have been designated as follows: (1) Personal Self Sufficiency, (2) Community Self Sufficiency, (3) Personal-Social Responsibility (all from Part I), (4) Personal Adjustment, and (5) Social Adjustment (Part II).

Broadly speaking, the purposes of the ABS:SE are for assessment of the child to help in instructional planning and in the development of individualized education programs. The manuals state that any adult who has had an opportunity to observe the child (e.g., teachers, parents, speech therapists, etc.) can act as an informant. The technical manual encourages the professional using the ABS:SE to use both teacher and parent data, where possible, when evaluating profiles of performance (Lambert, 1981). The ABS:SE can be administered in either of two ways; i.e., first-person assessment and third-party assessment. First-person assessment is used when the rater both is experienced with the scale and knows the child well. In such cases the person fills in the scale directly himself or herself. Third-party assessment is used when the informant is not sufficiently trained to complete the scale alone, and someone trained in administration systematically questions the informant about each item.

Critique

In keeping with the emphasis of this report, only Part II of the ABS:SE will be assessed. The psychometric characteristics of the ABS:SE Part II are summarized in Table 2 and Appendix B. The ABS:SE is relatively well standardized for use with teachers, with standardization data available for 6,500 children in Regular, EMR, and TMR classes. Considerable attention was paid to the effects of race/ethnicity, sex, and population density during the standardization process, and these seemed to have little undue influence on domain scores. One apparent weakness with the standardization data is the lack of information on children below age 7 years in EMR settings. This is due in large part to difficulties in identifying such children in the earliest years, although it would seem possible to include data on at least some 5- and 6-year-olds in EMR classes. Unfortunately, there are no standardization data for parent ratings for the ABS:SE. The technical manual reports a study showing no *statistically significant* differences between teacher and parent ratings for a group of 120 students (Lambert, 1981). However, it would seem that the presumption of no difference is a poor substitute for the availability of real data with this important group of raters.

No internal consistency data for individual domains could be located for the ABS:SE. However, alpha coefficients were available for the Part II factor scores. Internal consistency was generally excellent for the Social Adjustment dimension but poor-to-mediocre for the Personal Adjustment factor. No item-total correlations could be located, and there was an absence of interrater or test-retest reliability data in the technical manual (Lambert, 1981).

As noted, the assignment of items to domains was on an *a priori* basis, and the composition of the individual domains in Part II is difficult to defend on empirical grounds or in the context of a coherent theory. However, unlike its close relative, the ABS: Residential and Community Edition, the scoring scheme for the ABS:SE does allow two factor scores to be calculated for Part II. This clearly is an improvement, although the factor scores are probably much too broad to offer much diagnostic precision. Lambert and Hartsough (1981) reported a study showing that a Composite Score, derived from the five factor scores, can discriminate between Regular, EMR, and TMR students. This could be interpreted as evidence of criterion group validity, although it should be noted that this comparison also used the three factors from Part I. Hence, the contribution of the maladaptive behavior domains is impossible to assess. Finally, the technical manual reports that the two factors derived from Part II domains were correlated at low-to-moderate levels with achievement scores (Lambert, 1981). This provides a modicum of evidence for the congruent validity of the Part II domains.

In summary, the ABS:SE is well standardized for use by teachers, but there is a regrettable lack of standardization data for parent ratings. There are no internal consistency data for the most frequently used clinical measures (namely domain scores), although alpha coefficients have been calculated for the five derived factors. These coefficients range from poor to excellent. The technical manual contains no data on item-domain correlations or interrater and test-retest reliability. The taxonomic/factorial validity of the Part II domains seems difficult to verify at this time. However, the ABS:SE can be scored onto empirically derived factors, although the maladaptive factors may be too broad for many clinical applications. There is a modicum of criterion group and congruent validity with the instrument. All in all, it does not seem that the ABS:SE has received nearly the attention to its psychometric properties that has been paid to the ABS: Residential and Community Edition. Given the similarity of the two instruments, their psychometric characteristics are probably similar in many respects (although this cannot merely be assumed). Nevertheless, there is a disappointing lack of data on Part II of the ABS:SE, especially with respect to its validity.

References

Lambert, N. (1981). *AAMD Adaptive Behavior Scale–School Edition. Diagnostic and technical manual.* Austin, TX: Pro-Ed.

Lambert, N.M. & Hartsough, C.S. (1981). Development of a simplified diagnostic scoring method for the school version of the Adaptive Behavior Scale. *American Journal of Mental Deficiency, 86*, 138-147.

Lambert, N.M. & Nicoll, R.C. (1976). Dimensions of adaptive behavior of retarded and nonretarded public-school children. *American Journal of Mental Deficiency, 81*, 135-146.

Lambert, N., Windmiller, M., Tharinger, D. & Cole, L. (1981). *AAMD Adaptive Behavior Scale–School Edition. Administration and instructional planning manual.* Austin, TX: Pro-Ed.

Nihira, K. (1973). Importance of environmental demands in the measurement of adaptive behavior. In G. Targan, R.K. Eyman & C.E. Meyers (Eds.), *Sociological studies in mental retardation* (pp. 101-116). Monographs of American Association on Mental Deficiency #1.

Aberrant Behavior Checklist (ABC)

M. G. Aman & N. N. Singh, 1986

Point-form Synopsis

Stated Purpose: To assess the effects of pharmacological, behavioral, dietary, or other treatments that may have an impact on behavior. To assess inappropriate and maladaptive behavior in mentally retarded children and adults without respect to treatment.

Age Range: Scale developed on samples ranging from 5 years through adulthood.

Level of Mental Retardation Covered: Scale developed on samples with moderate through profound mental retardation.

Raters/Diagnosers: Personnel, such as unit supervisors, teachers, nurses and nurse aides, and other caretakers who have regular contact with the individual being rated.

Time Required to Complete: Approximately 5 to 7 minutes for the rating portions, with an additional 5 minutes if the (optional) face sheet is filled out.

Disorders/Dimensions Identified: Five subscales as follows: (1) Irritability, Agitation, Crying; (2) Lethargy, Social Withdrawal; (3) Stereotypic Behavior, (4) Hyperactivity, Noncompliance; and (5) Inappropriate Speech.

Date of Manual Publication: 1986.

Cost: ABC kit (manual plus 50 checklist forms and score sheets), $32.00. Package of 50 checklists and score sheets, $16.00.

Source: Slosson Educational Publications, Inc., P.O. Box 280, East Aurora, NY 14052. Telephone (716) 652-0930.

Limitations/Exclusions: Relevance to young children (≤ 10 years) uncertain due to small representation in developmental studies. Relevance to noncustodial settings uncertain.

Description

The Aberrant Behavior Checklist (ABC) is an informant based scale that originally was developed for use in treatment research, such as in studies of the effectiveness of psychotropic medication. It was derived by factor analysis and a cross validation procedure on two samples, totaling 927 individuals in residential institutions. The ABC comes with a face sheet which requests a variety of information, such as provision or not of specialized training, degree of mental retardation, current medical status, any medications taken, and so forth. Completion of the face sheet, especially after the first administration of the scale, often is unnecessary if serial ratings are to be obtained. The actual rating portion of the ABC has 58 behavioral items which describe maladaptive or inappropriate behavior. These items resolve into five subscales as follows: (1) Irritability, Agitation, Crying (15 items), (2) Lethargy, Social Withdrawal (16 items), (3) Stereotypic Behavior (7 items), (4) Hyperactivity, Noncompliance (16 items), and (5) Inappropriate Speech (4 items). Each item is described in more concrete terms in the manual. Higher scores on the ABC signify more serious inappropriate or maladaptive behavior. The rating portions of the checklist typically take about 5 minutes to complete. If information on the face sheet is needed, it may require an additional 5 minutes to fill in.

Although the checklist was developed to assess the effects of treatments, it also may be useful for identifying individuals in need of intervention or for selecting persons suitable for participation in scientific studies. Recently, articles have appeared in which the ABC was used either to select or to describe the subjects under investigation (Matson & Keyes, 1988; Sturmey, Carlsen, Crisp, & Newton, 1988).

Critique

Psychometric characteristics of the ABC are summarized in Table 2 and Appendix B. The ABC was developed in populations with moderate through profound mental retardation (Aman, Singh, Stewart, & Field, 1985a). More recently, the instrument has been assessed with a sample having borderline IQ and mild mental retardation without apparent loss to the scale's psychometric integrity (Rojahn & Helsel, 1989). The manual for the ABC presents average subscale scores and deviation units for large samples of residents in institutions in the United States and New Zealand. Average subscale scores and deviation units are not yet available for noninstitutionalized populations, although studies to do this are underway.

The internal consistency of the checklist has been found to be consistently very high across studies, with mean alpha levels ranging from .84 to .93 (Aman et al., 1985a; Bihm & Poindexter, in press; Freund & Reiss, 1990; Newton & Sturmey, 1988; Rojahn & Helsel, 1989). Likewise, item total correlations ranged from .39 to .88 (mean =.60), levels which on the average are very high.

Data on the scale's reliability are less clear-cut. Initially, extremely high test-retest reliability levels were reported for this scale (in the high .90s) (Aman, Singh, Stewart, & Field, 1985b), but these were later discounted by two of the original authors and another researcher (Aman, Singh, & Turbott, 1987). Instead, test-retest reliability appears to lie in the .70s, depending in part on the type of instructions, subscale being assessed, and rater effects. On the average, these correspond to adequate-to-good levels of agreement. Interrater reliability may be more problematic. Depending upon the raters used, instructions given, and subscale assessed, interrater reliability levels have averaged in the high .50s and low .60s (Aman et al., 1985b; Aman, Richmond, Stewart, Bell, & Kissel, 1987). These indicate acceptable, but not high, levels of agreement between raters. Rojahn and Helsel (1989) reported lower levels of interrater reliability (mean $r = .50$), but as noted by the authors themselves, no attempt was made to hold raters constant (i.e., to use the same raters to assess a group of subjects) in that study, and the time of observation was only 8 hours. Low reliabilities reported by Freund and Reiss (1990) employed ratings from different settings, which is known to depress rater agreement (Achenbach, McConaughy, & Howell, 1987).

There is a substantial body of data attesting to the checklist's validity. The instrument was developed with a New Zealand population, using factor analytic procedures and two large independent samples to cross-validate the initial factor structure. This factor structure was largely replicated in several studies conducted with United States (Aman et al., 1987; Bihm & Poindexter, in press) and English (Newton & Sturmey, 1988) residential populations. It also was replicated with a much younger sample (mean age = 10 years), that had a substantial representation of subjects with borderline IQs or mild retardation (Rojahn & Helsel, in press). Criterion group validity has been addressed in a number of ways. For example, subjects attending special educational facilities and those with Down syndrome were rated as having significantly lower scores than those unable to attend and subjects not having Down syndrome, respectively (Aman et al., 1985b). Likewise, subjects taking psychoactive medications and those with a diagnosis of psychosis obtained significantly and substantially higher ratings than unmedicated and nondiagnosed subjects, respectively (Aman et al., 1987). Rojahn and Helsel (in press) found that several subscales differentiated between diagnostic groups based on the DSM-III. For example, subjects with Organic Mental Syndromes and Infantile Autism scored particularly high on the Lethargy, Social Withdrawal subscale, and those diagnosed as having autism scored higher than all other groups on Stereotypic Behavior. Likewise, subjects who tested positive on the Dexamethasone Suppression Test (DST) had significantly higher scores on the Irritability, Agitation subscale than DST suppressors, even though psychiatric evaluation failed to differentiate between the two groups (Raft & Richmond, 1989). Congruent validity has been assessed by comparing ABC scores to those from other behavior rating instruments and by direct observation of behavior categories similar to those addressed in the checklist (Aman et al., 1985b). ABC subscales were found to correlate negatively with adaptive behavior as assessed on several instruments, positively with their respective counterparts from Part II (maladaptive behavior) of the AAMD Adaptive Behavior Scale (convergent validity) (Nihira, Foster, Shellhaas, & Leland, 1974), and not at all with IQ scores (divergent validity). All except one subscale were correlated with analogous categories assessed by behavior observations. In another study (Sturmey & Ley, 1990), several subscales from the ABC were significantly and substantially correlated with analogous subscales on the Psychopathology Instrument for Mentally Retarded Adults (Matson, 1988).

To summarize, there are substantial data available for the ABC on average subscale scores and standard deviation units for institutionalized subjects who are moderately to profoundly retarded. However, at the time of this writing, such data are not available for mildly retarded individuals and persons residing in the community. Internal consistency appears to be good, with alpha coefficients averaging about .90 across subscales and studies. Data on the checklist's reliability have been mixed. Test-retest reliability generally has been rated as adequate to good, but interrater reliability typically has been lower and appears to fluctuate with a variety of factors, such as subscale assessed, instructions provided, and so forth. Factorial validity for the ABC appears to be well established, the original factor structure having been replicated in different countries and with samples having quite different compositions. Several comparisons have attested to the criterion group validity of the instrument. Although the ABC was not developed as a diagnostic tool, some data exist , however, suggesting that subscale scores may be related to DSM-III diagnoses and DST results. Congruent validity has been determined by moderate relationships in the expected direction with adaptive behavior, maladaptive scales, and direct observations.

In conclusion, the ABC generally stands up well psychometrically, although interrater reliability may not be as satisfactory as desired and may be worthy of more research. The scale has been used extensively in drug research and has proven to be quite sensitive for that purpose (Aman & White, 1986). The ABC was not developed for use as a screening or diagnostic instrument, although that does not preclude the scale's use for that purpose. The ABC may prove to be an acceptable tool for subject selection or other identification purposes, but more research that specifically assesses its usefulness for that purpose would be desirable before adopting the checklist to that end.

References

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross informant correlation for institutional specificity. *Psychological Bulletin, 101*, 213-232.

Aman, M.G., Richmond, G., Stewart, A.W., Bell, J.C., & Kissel, R.C. (1987). The Aberrant Behavior Checklist: Factor structure and the effect of subject variables in American and New Zealand facilities. *American Journal of Mental Deficiency, 91*, 570-578.

Aman, M.G. & Singh, N.N. (1986). *Aberrant Behavior Checklist: Manual.* East Aurora, NY: Slosson Educational Publications.

Aman, M.G., Singh, N.N., Stewart, A.W., & Field, C.J. (1985a). The Aberrant Behavior Checklist: A behavior rating scale for the assessment of treatment effects. *American Journal of Mental Deficiency, 89*, 485-491.

Aman, M.G., Singh, N.N., Stewart, A.W., & Field, C.J. (1985b). Psychometric characteristics of the Aberrant Behavior Checklist. *American Journal of Mental Deficiency, 89*, 492-502.

Aman, M.G., Singh, N.N., & Turbott, S.H (1987). Reliability of the Aberrant Behavior Checklist and the effect of variations in instructions. *American Journal of Mental Deficiency, 92*, 237-240.

Aman, M. G., & White, A. J. (1986). Measures of drug change in mental retardation. In K. D. Gadow (Ed.), *Advances in learning and behavioral disabilities* (Vol. 5, pp. 157-202). Greenwich, CT: JAI Press.

Bihm, E., & Poindexter, A. R. (in press). Cross-validation of the factor structure of the Aberrant Behavior Checklist for mentally retarded persons. *American Journal on Mental Retardation.*

Freund, L. S., & Reiss, A. L. (1990). *Rating psychopathology in mentally retarded outpatients: Use of the Aberrant Behavior Checklist.* Unpublished manuscript, Department of Psychiatry, Johns Hopkins University.

Matson, J. L. (1988). *The PIMRA manual.* Orland Park, IL: International Diagnostic Systems, Inc.

Matson, J.L. & Keyes, J. (1988). Contingent reinforcement and contingent restraint to treat severe aggression and self-injury in mentally retarded and autistic adults. *Journal of the Multihandicapped Person, 1*, 141-153.

Newton, J.T. & Sturmey, P. (1988). The Aberrant Behavior Checklist: A British replication and extension of its psychometric properties. *Journal of Mental Deficiency Research, 32*, 87-92.

Nihira, K., Foster, R., Shellhaas, M., & Leland, H. (1974). *AAMD Adaptive Behavior Scale* (rev. ed.). Washington, DC: American Association on Mental Deficiency.

Raft, E.V., & Richmond, G.H. (1989). *Results of the dexamethasone suppression test with profoundly retarded institutionalized patients.* Unpublished manuscript, Murdoch Center, Butner, NC.

Rojahn, J. & Helsel, W.J. (in press). The Aberrant Behavior Checklist in children and adolescents with dual diagnosis *Journal of Autism and Developmental Disorders.*

Sturmey, P., Carlsen, A., Crisp, A.G., & Newton, J.T. (1988). A functional analysis of multiple aberrant responses: A refinement and extension of Iwata et al.'s (1982) methodology. *Journal of Mental Deficiency Research, 32*, 31-46.

Sturmey, P., & Ley, T. (1990). The Psychopathology Instrument for Mentally Retarded Adults: Internal consistencies and relationship to behaviour problems. *British Journal of Psychiatry, 156*, 428-430.

Adolescent Behavior Checklist
H. B. Demb, N. Brier, & R. Huron, 1989

Point-form Synopsis

Stated Purpose: To identify individuals, aged 12 to 21 years, who are at high risk of having a diagnosable psychiatric disturbance.

Age Range: Twelve to 21 years.

Level of Mental Retardation Covered: Borderline intelligence and "high" mild mental retardation.

Raters/Diagnosers: Individuals with borderline and mild mental retardation able to understand and respond to scale items. Administration guided by suitable adult.

Time Required to Complete: Reported as 20 to 30 minutes. Estimated by reviewer at 15 to 25 minutes.

Disorders/Dimensions Identified: Eight diagnostic groupings as follows: (1) Anxiety, (2) Hyperactivity, Impulsivity, Inattention, (3) Conduct Disorder, (4) Oppositional Disorder, (5) Affective Illness, (6) Psychosis, Autistic, Schizoid, (7) Intake/Control, and (8) Trait Disorder. In addition, there is a Lie Scale, a Total Score, and a Clinical Score (Total minus the Lie Score).

Date of Manual Publication: No manual available. The revised rating form is dated 1989.

Cost: Unknown. Not commercially available.

Source: Howard Demb, M.D., Albert Einstein College of Medicine at Yeshiva University, Children's Evaluation and Rehabilitation Center, Rose F. Kennedy Center, 1410 Pelham Parkway South, Bronx, NY 10461. Telephone (212) 430-2443 and 430-2441.

Limitations/Exclusions: Not appropriate for individuals younger than 12 or older than 21 years. Not suitable for persons with lower levels of mild mental retardation or with moderate, severe, or profound mental retardation.

Description

The Adolescent Behavior Checklist is a self-report screening instrument designed to identify youngsters between the ages of 12 and 21 years who are at risk of having a diagnosable mental illness. The Checklist uses DSM-III-R criteria, and it was developed for use with adolescents having borderline intellectual functioning or those falling into the higher levels of mild mental retardation. The Checklist is made up of 86 items which render scores on eight subscales derived from the DSM-III-R. The eight subscales are designated as follows: (1) Anxiety (14 items), (2) Hyperactivity, Impulsivity, Inattention (8 items), (3) Conduct Disorder (8 items), (4) Oppositional Disorder (7 items), (5) Affective Illness (12 items), (6) Psychosis, Autistic, Schizoid (9 items), (7) Intake/Control (10 items), and (8) Trait Disorder (12 items). In addition, there is a Lie Scale (6 items) to enable an estimate to be made about how honest the subject was while responding to the scale. Higher scores on each of these subscales indicate the likelihood of a more serious disorder or behavior problem. A Total score is derived by adding affirmative responses for all 86 items. Finally, a Clinical score is calculated by subtracting the Lie score from the Total score.

Cosgrove-Dapuzzo (1989) suggests three other possible applications for the checklist in addition to its function of screening subjects for mental disorders. These include the following: (1) Identification of ways in which the adolescents' views of themselves differ from what others report, (2) Selection of specific problems requiring therapeutic change, and (3) Documentation of eligibility for services to the emotionally disturbed.

The items within the Adolescent Behavior Checklist use language designed for subjects with a fourth-grade listening level (Cosgrove-Dapuzzo, 1989). The instructions call for the items to be read to the person, who is asked to reply yes or no to each question. Each yes response signifies the presence of the given symptom and is given a weight of 1 in its respective diagnostic group (subscale). There is no overlap between items; that is, each item is scored onto one, and only one, diagnostic group. The writer estimates that it would take approximately 15 to 25 minutes to administer the scale, depending in large part upon the subject's language ability, understanding of the terms, and cooperativeness. It takes approximately 5 minutes to score the checklist (Cosgrove-Dapuzzo, 1989).

Instructions for the Adolescent Behavior Checklist do not specify who is qualified to administer the questionnaire. However, it would appear that any caretaker or professional with the capability of establishing good rapport with the adolescent and the ability to follow the standard procedures outlined in the instructions would be able to administer the instrument satisfactorily.

Critique

The psychometric characteristics of the Adolescent Behavior Checklist are summarized in Table 2 and Appendix B. The checklist is a very new instrument, and yet there is a remarkable amount of relevant psychometric information available . An appendix to the instructions provides suggested cutoff scores for each symptom category for identifying adolescents who are likely to have diagnosable psychiatric disturbances (Demb, Brier, & Huron, 1989). However, no standardization data are provided about the samples, such as the number of adolescents in each group, their ages, IQ scores, sex ratios, and other relevant data that would enable the user to judge whether these samples are representative of adolescents having borderline/mild mental retardation. Hence, these cutoffs would have to be treated as very tentative at the present time.

The internal consistency of the checklist ranges from mediocre to good. The alpha coefficients extended from .58 (for the Intake/Control subscale) to .91 (Oppositional subscale), with a mean of .76 for the eight diagnostic categories (Cosgrove-Dapuzzo, 1989). Six of the eight subscales had alpha coefficients greater than .70, a level regarded by many to be acceptable (Reiss, 1988). However, the alpha value for the Lie subscale was only .25. It is difficult to know whether this is due to a lack of endorsement (i.e., perhaps most respondents simply do not falsify) or to a weakness in the subscale itself, but it is clear that this issue needs to be researched further before Lie scores can be taken at face value. Test-retest reliability was assessed over a 3-week period and found to be very high (Cosgrove-Dapuzzo, 1989). For example, for a combined group of 40 subjects, test-retest reliability was found to range from .87 to 1.00, depending upon the subscale assessed, with a mean test-retest reliability of .96 across all eight subscales (Cosgrove-Dapuzzo, 1989).

The factorial/taxonomic validity of the Adolescent Behavior Checklist depends in large part on the validity of the DSM-III-R itself, as its items were adapted from the latter. However, the checklist is on stronger ground than many other instruments which have used existing psychiatric schemes as their source because its application is intended to be confined to adolescents with borderline intelligence and those with high mild mental retardation. As pointed out in the Introduction to this review, the application of traditional diagnostic approaches becomes increasingly questionable as the severity of mental retardation increases. Criterion validity has been addressed by comparing 20 adolescents having an emotional disturbance (diagnosed in accordance with the DSM-III-R) with 20 control subjects. The emotionally disturbed group scored significantly higher than the controls on seven of eight subscales of the checklist as well as on the Total score (Cosgrove-Dapuzzo, 1989). However, no attempt was reported to determine whether subjects having specific DSM diagnoses presented with higher scores on analogous subscales of the Adolescent Behavior Checklist. Congruent validity was addressed by having the subjects rate themselves on the Youth Self-Report Form (Achenbach & Edelbrock, 1987) and by having teachers rate them on the Teacher Report Form (Achenbach & Edelbrock, 1986). By and large, there was a moderately good relationship between the checklist and the Youth Self-Report. The Anxiety, Affective Illness, and Trait Disorder subscales of the former were correlated with an Internal Problem domain of the latter. Likewise, Hyperactivity, Conduct Disorder, and Oppositional subscales of the checklist were all correlated with the Externalizing domain from the Youth Self-Report. However, only a minority of subscales from the Checklist were related significantly to their analogous subscales on the Youth Self-Report. Ratings on the Teacher Report Form were related weakly to checklist scores. The Total scores of both scales were moderately correlated (r=.56), but there were few significant correlations between congruent subscales on the two instruments.

To summarize, the characteristics of the standardization samples for the Adolescent Behavior Checklist are largely unknown at the present. The internal consistency of its subscales, with a mean alpha of .76, appears to be satisfactory. However, as previously stated, the Lie scale has an unsatisfactory level of internal consistency (alpha=.25) and, therefore, requires more research. The test-retest reliability of the instrument appears to be extremely high (mean r=.96). Taxonomic validity of the checklist must be inferred from its relationship to the DSM-III-R and the appropriateness of the latter to the mentally retarded population. Criterion group validity appears to be established, at least for persons with certain disorders, provided that one is not interested in screening for specific types of disorder. Likewise, the congruent validity of the checklist is supported for the Total score, especially when another self-report scale is used as the criterion. However, validity breaks down when the correspondence between individual subscales is considered. One might argue that this is a moot point, as the checklist reportedly is intended only as a general screening instrument. Nevertheless, by virtue of the fact that it contains eight distinct subscales, each with its own cutoff value, some users inevitably will be tempted to employ it in a more specific manner to derive inferences about particular types of behavior problems.

To conclude, the Adolescent Behavior Checklist appears to be a promising instrument, although it is still in the early stages of development. Its taxonomic validity hinges on that of the DSM-III-R, but to the extent that its use is confined to subjects with "high" mild mental retardation and borderline intelligence, this extension of DSM criteria appears to be defensible. Insofar as the checklist is used as a screening instrument, it appears to have promise. It also may have value as a standard inventory to explore self-appraisals, but the data do not support its use for differential diagnosis.

References

Achenbach, T.M,. & Edelbrock, C.S (1986). *Manual for the Teacher's Report Form and Teacher Version of the Child Behavior Profile.* University of Vermont, Burlington, VT: Authors.

Achenbach, T.M., & Edelbrock, C.S. (1987). *Manual for the Youth Self-Report.* University of Vermont, Burlington, VT: Authors.

Cosgrove-Dapuzzo, D. (1989). *An adolescent behavior checklist for identifying youngsters who are at high risk of having a diagnosable psychiatric disturbance: A pilot study.* Unpublished Doctor of Philosophy dissertation, Yeshiva University, New York, NY.

Demb, H., Brier, N., & Huron, R. (1989). *Adolescent Behavior Checklist (ABC).* Unpublished instrument available from Howard Demb, M.D., Albert Einstein College of Medicine at Yeshiva University, Bronx, NY.

Reiss, S. (1988). *Test manual for the Reiss Screen for Maladaptive Behavior.* Orland Park, IL: International Diagnostic Systems.

Balthazar Scales of Adaptive Behavior
II. Scales of Social Adaptation
E. E. Balthazar, 1973

Point-form Synopsis

Stated Purpose: To measure the effects of treatment, training, and other types of programs for individuals in residential institutions, day care centers, and clinics.

Age Range: Not specified. Scale developed on sample aged 5 years through adulthood.

Level of Mental Retardation Covered: Primarily severe and profound.

Raters/Diagnosers: Observers trained to record 25 maladaptive behaviors.

Time Required to Complete: Typically six 10-minute sessions.

Disorders/Dimensions Identified: Seven maladaptive subscales as follows: (1) Failure to Respond, (2) Stereotypy, Posturing, Including Objects, (3) Non-Directed, Repetitious Verbalization, (4) Inappropriate Self-Directed Behavior, (5) Disorderly, Non-Social Behavior, (6) Inappropriate Contact with Others, and (7) Aggressive Withdrawal.

Date of Manual Publication: 1973.

Cost: Manual, $5.50; specimen set, $7.00; complete kit (manual and materials for 25 ratings), $23.00; tally sheets (pad of 50), $6.50; scoring summary sheets (package of 25), $6.00.

Source: Consulting Psychologists Press, 577 College Avenue, Palo Alto, CA 94306-1490. Telephone (415) 857-1444.

Limitations/Exclusions: Relevance to mildly and moderately retarded as well as nonambulatory retarded persons uncertain.

Description

Unique among the instruments reviewed thus far is the Balthazar Scales of Adaptive Behavior: II. Scales of Social Adaptation (BSAB-II), because the scales are dependent upon timed direct observations of the individual for scoring purposes. The Balthazar Scales were developed by factor analysis of 71 behaviors that were observed directly in dayrooms and play yards of institutionalized residents, most of whom had severe or profound mental retardation. The original empirical publication of the Balthazar Scales reported an 18-factor solution that encompassed both adaptive behaviors and maladaptive behavioral categories (Balthazar & English, 1969). The behavioral groupings and subscales were subsequently modified to produce 19 categories (subsuming some 74 items), which formed the definitive version of the Balthazar Scales (Balthazar, 1973). Unfortunately, the manual does not clarify how the discrepancy occurred between the original number of items and factor solution reported by Balthazar and English (1969) and those adapted in the definitive scale.

In keeping with the focus of this review on behavior and psychiatric disorders, only the "Unadaptive Self-Directed" subscales will be described here in any detail. The seven unadaptive dimensions were labeled as follows: (1) Failure to Respond (4 items), (2) Stereotypy, Posturing, Including Objects (7 items), (3) Non-directed Repetitious Verbalization; Smiling, Laughing Behaviors (3 items), (4) Inappropriate Self-Directed Behavior (2 items), (5) Disorderly, Non-Social Behavior (3 items), (6) Inappropriate Contact with Others (2 items), and (7) Aggression, Withdrawal (4 items). In addition, there are 12 adaptive categories, collectively encompassing 49 items.

The manual states that "any articulate person who is conscientious, alert, and accurate" may administer and score the Balthazar Scales. The instrument was developed within an institutional milieu, and the instructions call for the person to be observed in the dayroom or play yard of a clinic, day care, or residential center. In general, it is recommended that the rater observe the subject in 10 one-minute units when performing ratings. For many categories, a partial-interval recording method is adopted, in which the behavior is recorded for occurrence or non-occurrence within the one-minute interval (e.g., subscales #2, 3, 4, and 5). For the remaining three subscales, the actual number of occurrences of the defined behavior is counted or tallied within each one-minute interval. The manual suggests obtaining several 10-minute samples at varied and representative times of the day spread over the week. The manual also suggests that six 10-minute sessions often have sufficed for descriptive purposes, although the scale's developers have used up to 12 sessions for some individuals.

The stated purposes of the BSAB-II are for measuring the effects of treatment, training, and other types of programs. The manual also suggests that the Scales of Social Adaptation may be helpful in developing training programs. As noted, the BSAB-II was developed in an institutional context with subjects mostly in the severely and profoundly mentally retarded range. The manual does not indicate specifically what populations may be assessed with the scales, although it does state that the instrument can be employed in residential institutions, day care centers, clinics, and foster homes and that any conscientious person, including parents, may be trained to use the scales.

Critique

Only the maladaption subscales are reviewed here; their psychometric characteristics are summarized in Table 2 and Appendix B. Average subscale scores for 100 institutionalized residents are presented in the manual (Balthazar, Rocca, & Rifkin, 1971). These scores seem to be too narrowly based for many clinical comparisons today, as such applications are increasingly likely to occur in community, rather than residential, settings. Furthermore, few data are provided about the reference group, such as the IQ levels subsumed, ages, gender, or the nature of the setting.

No data are provided about the internal consistency or test-retest reliability of the BSAB-II maladaptive subscales. Interrater reliability was addressed in two studies, and the proportion agreement ranged from .58 to .76 in one case and from .75 to .97 in the other. The reviewer finds it difficult to believe that levels this high could be obtained consistently with such a complex instrument. Note that the unit of measurement is not the 19 subscales of the BSAB-II but, instead, the 74 behavioral items encompassed within the subscales. Most investigators find that reliability begins to deteriorate as the number of categories is increased from a workable number (e.g., six or eight) to a large number (e.g., 10 or 12) (Aman & White, 1986). The 74 items of the BSAB-II would seem to tax even the most conscientious of observers.

The factorial validity of the scales was addressed in the Balthazar and English (1969) study in which factor structure of the entire instrument was determined. Unfortunately, the definitive scales were only derived "for the most part" from the earlier study (Balthazar, 1973, pg. 4), and the final instrument contains some significant departures from the structure reported in the Balthazar and English study. No data could be located that addressed criterion group or congruent validity of the BSAB-II maladaptive behavior subscales. Balthazar argues in the manual that it is not relevant to ask if observer judgments are valid, as the behavioral occurrences are considered to have validity in and of themselves. However, it seems to the reviewer that this hard-line behavioral position really begs the question. What surely is at issue is whether or not the categories covered in the scales relate to clinically meaningful dimensions. No external data could be located to substantiate this.

Two studies addressed validity of a different type; namely, behavioral changes in response to therapeutic programs. In one study, subject responses to staff nurturance plus the use of behavior modification principles were evaluated (Balthazar, English, & Sindberg, 1971), whereas the effects of enhanced stimulation plus supplementary activities emphasizing acquisition of self-help skills were measured in the other (Naor & Balthazar, 1974). Both studies indicated significant changes on about half of the maladaptive behavior subscales assessed. Most of these changes indicated behavioral improvements, although some appeared to reflect worsening, which was probably an artifact of greater staff-resident contact. For example, there were more negative staff-subject interactions because, in fact, far more interactions took place (Balthazar, English, & Sindberg, 1971). Thus, available evidence suggests that the BSAB-II is apparently sensitive to the effects of therapeutic programs.

In summary, the standardization data for the BSAB-II are rather limited, both in terms of the number of subjects involved and also in terms of diversity (e.g., normative data are provided only for the institutional context). No data are available in relation to internal consistency or test-retest reliability. Although adequate reliability levels are reported in the manual, the complexity of the instrument is such that most observers probably would struggle to approach these standards. The factorial validity of the scales is uncertain due to alterations in the structure of the instrument after its original development. Criterion group and congruent validity were not addressed in the manual or other sources that the reviewer checked. Given the complexity of this tool, it would seem to require highly trained observers and a skilled professional to train and monitor observers. The scales do appear to be sensitive to various forms of behavioral, ecological, and training programs. The BSAB-II is historically important because it is perhaps the only available diagnostic device based solely on direct observation. As such, it serves as a unique source of information regarding the structure of maladaptive behavior in persons with severe and profound mental retardation. However, the instrument is rather unwieldy and, if the same standards that are used with rating scales are applied to the BSAB-II, it appears to fall somewhat short.

References

Aman, M. G., & White, A. J. (1986). Measures of drug change in mental retardation. In K. D. Gadow (Ed.), *Advances in learning and behavioral disabilities* (Vol. 5, pp. 157-202). Greenwich, CT: JAI Press.

Balthazar, E. E. (1973). *Balthazar Scales of Adaptive Behavior: II. Scales of Social Adaptation.* Palo Alto, CA: Consulting Psychologists Press, Inc.

Balthazar, E. E., & English, G. E. (1969). A factorial study of unstructured ward behaviors. *American Journal of Mental Deficiency, 74*, 353-360.

Balthazar, E. E., English, G. E., & Sindberg, R. M. (1971). Behavior changes in mentally retarded children following the initiation of an experimental nursing program. *Nursing Research, 20*, 69-74.

Balthazar, E. E., Rocca, A., & Rifkin, H. (1971). *Data on one hundred retardates in a residential institution.* Madison, WI: Central Wisconsin Colony and Training School Research Department.

Naor, E. M., & Balthazar, E. E. (1974). The effects of a self-help training programme upon social coping behaviours. *British Journal of Mental Subnormality, 20*, 69-77.

Behaviour Disturbance Scale
I. Leudar, W. Fraser, & M. A. Jeeves, 1984

Point-form Synopsis

Stated Purpose: To assess behavior disturbances, plan rehabilitation, and carry out research on relocation, treatment, and age-related changes (Hogg & Raynes, 1987).

Age Range: Individuals 15 years of age and older.

Level of Mental Retardation Covered: Mild through severe.

Raters/Diagnosers: Training center instructors, nurses, parents, and other pertinent caretakers with knowledge of the person.

Time Required to Complete: Reported to be approximately 15 minutes. (Estimated by reviewer to be less than 10 minutes.)

Disorders/Dimensions Identified: Six subscale scores as follows: (1) Aggressive Conduct, (2) Mood Disturbance, (3) Communicativeness, (4)Antisocial Conduct, (5) Idiosyncratic Mannerisms, and (6) Self Injury.

Date of Manual: 1985.

Cost: A once only "cost of materials fee" is charged: £10 for students and £30 for all other workers. Purchasers are free to make personal copies of all materials thereafter.

Source: Dr. Ivan Leudar, Psychology Department, The University of Manchester, Manchester M13 9PL, England.

Limitations/Exclusions: Not appropriate for children and adolescents less than 15 years of age. Not suitable for profoundly retarded individuals (i.e., those with no language).

Description

The Behaviour Disturbance Scale (BDS) is a 51-item checklist, developed for assessing problem behaviors in mentally retarded adults. Each item of the instrument is scored with a 5-point scale: (1) *never* through to (5) *very frequently* . During the development of the BDS, nurses working in residential centers and training center instructors conducted all ratings, but any caretaker with a good knowledge of the person presumably could perform such ratings. Higher scores on most subscales signify more serious maladaptive behavior, although higher scores on one subscale, Communicativeness, reflect more adaptive behavior.

The Behaviour Disturbance Scale was developed by factor analysis of the ratings of mentally retarded adults. In the first stage of the scale's development (BDS 1), 629 individuals, ranging in age from 16 to 45 years, were rated by caretakers on a series of 20 behavioral items. A principal components analysis with varimax rotation was used, and six factors were extracted that accounted for 62% of the variance. The factors were labeled as follows: (1) Aggressive Conduct, (2) Mood Disturbance, (3) Idiosyncratic Mannerisms, (4) Communicativeness, (5) Overactivity, and (6) Antisocial Conduct. In the second stage of the scale's development (BDS 2), the list of behavioral symptoms was increased from 20 to 51 items. Two hundred forty-seven adults, residing in institutions and in the community, were rated by caretakers who had known them for at least 6 months. The results were factor analyzed, using the statistical methods employed for the BDS 1, and the outcome was six factors, five of which were similar to those in the previous analysis. The six factors, making up the definitive version of the Behaviour Disturbance Scale, are designated as follows: (1) Aggressive Conduct (12 items), (2) Mood Disturbance (11 items), (3) Communicativeness (11 items), (4) Antisocial Conduct (7 items), (5) Idiosyncratic Mannerisms (9 items), and (6) Self Injury (4 items). The first five subscales are essentially the same as in the derivation of BDS 1, whereas Self Injury appeared anew in this analysis.

The method for scoring the BDS is not stated in the publication reporting its development (Leudar, Fraser, & Jeeves, 1984). However, in a subsequent paper, Leudar and Fraser (1987) report a procedure for weighting each item for degree of seriousness. The weighting procedure and the methods for developing the weighting procedure were not described in sufficient detail, however, to summarize them here.

Additional Features

A computerized version of the BDS is available from the scale's developers. This allows for direct entry of ratings on a microcomputer and provides summary scores when entry is completed.

Critique

Research on psychometric characteristics of the BDS is summarized in Table 2 and Appendix B. Unfortunately, normative data (average subscale scores) have not been published for the BDS, which makes it difficult to interpret individual profiles using the scale. Also, the size of the sample (247) used to derive the factor structure of the definitive scale was somewhat small, although the available evidence suggests that the factor structure is quite robust.

To the best of the writer's knowledge, no measure of the instrument's internal consistency has been presented, such as alpha, Spearman-Brown, or item-whole correlations. The writer was unable to locate a measure of test-retest reliability per se, although one report used initial ratings to predict subsequent ratings two years later (Leudar et al., 1984). Original subscale scores or their transformations predicted between 24% and 58% of outcome variance. However, consistency for one subscale, Self Injury, was not reported. There are relatively few data on the interrater reliability of the BDS. In one exercise, interrater reliability was assessed for 10 subjects and a correlation of .75 was obtained, but this presumably was confined to the Total Score measure. A subsequent assessment of interrater reliability on 16 subjects found a range of correlations from .65 to .89, corresponding to good-to-excellent agreements (Leudar et al., 1984). A third exercise reported a correlation of .89 but, again, this was presumably only for the Total Score, and the sample size was not specified.

Factorial validity has been established through the instrument's two-stage development using factor analysis. Analysis of a predecessor of the BDS (i.e., BDS 1) resulted in five factors similar to five dimensions in the definitive scale. Expansion of the instrument to 51 items resulted in the current six subscales, which were factorially derived. Criterion group validity has been assessed by comparing ratings of institutionalized residents with persons living in the community. Institutional residents received significantly higher scores on the Aggressive Conduct, the Antisocial Conduct, and the Self Injury subscales (Leudar et al., 1984). In a subsequent study (Fraser, Leudar, Gray, & Campbell, 1986), similar relationships were found between certain BDS subscales and institutionalization. Congruent validity has been addressed by comparing BDS subscale scores with factors derived from psychiatrists' ratings using the Clinical Interview Schedule (Goldberg, Cooper, Eastwood, Kedward, & Shepherd, 1970). Although there were no strong relationships, the Communicativeness and Aggressive Conduct subscales were associated significantly with certain psychiatrically derived dimensions (Fraser et al., 1986).

To conclude, then, the major drawbacks of the BDS appear to be as follows. First, normative (or average subscale) data do not appear to be available in published form. Second, there is an absence of internal consistency data at this stage. Third, the data regarding the test-retest and interrater reliability of the BDS are relatively sparse. More congruent validity data also would be helpful in establishing what the individual subscales mean, as users of the instrument no doubt will want to make specific research and clinical inferences about individuals from the subscale profiles. On the positive side, it should be noted that the BDS was developed exclusively from ratings on persons with mental retardation whose functional deficits nearly cover the entire range of mental retardation. Furthermore, the factor structure of this instrument appears to be robust and to make clinical sense, especially in light of other factor analytic research with mentally retarded populations. To sum up, the BDS appears to be one of the more promising behavior rating scales for use in this field. However, additional research still is needed to deal with some of the questions raised above.

References

Fraser, W. I., Leudar, I., Gray, J., & Campbell, I. (1986). Psychiatric and behaviour disturbance in mental handicap. *Journal of Mental Deficiency Research*, 30, 49-57.

Goldberg, D. P., Cooper, B., Eastwood, M. R., Kedward, H. B., & Shepherd, M. (1970). A standardized psychiatric interview for use in community surveys. *British Journal of Preventive and Social Medicine*, *24*, 18-23.

Hogg, J., & Raynes, N. V. (Eds.) (1987). *A guide to assessment practices, tests, and checklists*. London: Croom Hlem.

Leudar, I., & Fraser, W. I. (1987). Behaviour disturbance and its assessment. In J. Hogg & N. V. Raynes (Eds.), *Assessment in mental handicap: A guide to assessment practices, tests, and checklists* (pp. 107-128). London: Croom Helm.

Leudar, I., Fraser, W. I., & Jeeves, M. A. (1984). Behaviour disturbance and mental handicap: Typology and longitudinal trends. *Psychological Medicine*, *14*, 923-935.

Client Development Evaluation Report (CDER) Department of Developmental Services, State of California, 1986

Point-form Synopsis

Stated Purpose: To assist interdisciplinary teams in assessing the developmental and emotional status of clients with developmental disabilities, and for determining service needs at the management level.

Age Range: Childhood through adulthood.

Level of Mental Retardation Covered: Mild through profound.

Raters/Diagnosers: Persons who interact with the individual on a regular basis.

Time Required to Complete: Not reported. Estimated by reviewer at 6 to 9 minutes for the Emotional Domain.

Disorders/Dimensions Identified: Fifteen behavior problems are rated on the Emotional Domain.

Date of Manual Publication: 1986.

Cost: No charge.

Source: Mr. James White, Department of Developmental Services, 1600 9th Street, Sacramento, CA 95814. Telephone (916) 323-7701; FAX (916) 323-4929.

Limitations/Exclusions: None identified.

Description

The Client Development Evaluation Report (CDER [pronounced same as cedar]) is an assessment instrument developed by the California Department of Developmental

Services (1986). It has two primary purposes; namely, (1) to collect data on client diagnostic characteristics and (2) to measure and evaluate the functioning levels of persons with developmental disabilities who receive services in the California developmental disabilities service system.

The CDER is made up of two principal components, the Diagnostic Element and the Evaluation Element. The Diagnostic Element uses information provided primarily by the individual's physician and psychologist. It contains a summary of the types, etiologies, and levels of severity of primary disabilities of the person and the likely impact of these on programming. Information collected includes the following: (1) etiology of the mental retardation, (2) presence of cerebral palsy, (3) existence of epilepsy/seizure disorder, (4) presence of other developmental disability, (5) presence of known risk factors, (6) any coexisting mental disorders, (7) major chronic medical conditions, (8) sensory acuity, (9) use of psychotropic drugs, (10) presence of abnormal involuntary movements, (11) special health requirements, (12) presence of serious behavior problems that might interfere with placement decisions, and (13) special legal conditions or constraints.

The Evaluation Element comprises 66 items and is used for recording the client's level of functioning. The items load onto six areas of development as follows: (1) Motor Domain, (2) Independent Living Domain, (3) Social Domain, (4) Emotional Domain, (5) Cognitive Domain, and (6) Communication Domain. The Evaluation Element can be completed by any responsible individual who interacts with the client on a regular basis. The Emotional Domain, which is the only one that will be reviewed here, is made up of 15 items. Seven of the items are scored in a 4-point ordinal scale, seven on a 5-point scale, and one on a 7-point scale. The *CDER Field Manual* does not state how individual domains are scored, as this is conducted centrally by computer. Presumably, items are simply totaled, and higher scores on the Emotional Domain appear to signify *fewer* behavior/emotional items. The composition of the Emotional Domain is heterogeneous in the sense that several different types of behavior problems (e.g., aggression, depression, stereotypy, wandering away) are encompassed within it.

The adaptive behavior items for the CDER were modeled closely on items from existing adaptive behavior scales, such as the AAMD Adaptive Behavior Scale (Widaman, Gibbs, & Geary, 1987). Its primary use is as a management tool for the State of California in the following ways: (1) calculating the number of persons with developmental disabilities, (2) serving budgetary functions, such as determining staffing requirements, (3) establishing priorities by assessing current needs, and (4) serving as a data base for aggregated reports (Department of Developmental Services, 1986). In addition, the field manual indicates that the CDER has several uses at the local level such as (1) determining appropriate placement, (2) monitoring program effectiveness, (3) planning prevention strategies, and (4) assessing future resource needs.

Additional Features

When scored centrally, a computer-generated Client Summary Profile is prepared for each person. In addition to the summaries of diagnostic and adaptive behavior, this produces a bar graph of maladaptive behavior, which is subdivided into intrapunitive and extrapunitive and which shows percentiles. Also, a score denoting total severity of maladaptive behavior is produced, which uses weights from a factor analysis of the CDER (Widaman et al., 1987), discussed below.

Critique

The available data regarding the CDER's psychometric characteristics are summarized in Table 2 and Appendix B. The field manual presents no normative or standardization data for the instrument. However, it does state that the CDER data system is the largest and most comprehensive in the world, and extensive data of this type presumably are stored centrally.

The reviewer was unable to locate data on the instrument's internal consistency or test-retest reliability. Interrater reliability levels for the Emotional Domain were reported to range from .60 to .90 for all except three items, which fell below .50. The mean overall reliability was not reported for the domain, but it would appear to be acceptable from these data (Harris, Eyman, & Mayeda, 1982).

Widaman et al. (1987) studied the factor structure of the CDER with a sample of over 6,000 subjects. The six-factor solution that was adopted included two factors related to behavior/emotional problems: Social (Extrapunitive) Maladaption and Personal (Intrapunitive) Maladaption. Eight items fell on the Social Maladaption factor and seven on the Personal Maladaption factor. However, two items from the Emotional Domain, namely Hyperactivity and Adjustment to Changes in the Environment, were not included, and it included one item (Unacceptable Social Behavior) that does not appear on the Emotional Domain. Nihira, Price-Williams, and White (1988) found that individuals who were diagnosed as having one of the five most common psychiatric disorders in the developmentally handicapped population served by California's Department of Developmental Services consistently were rated significantly lower on Social Maladaptation (i.e., as having more problems), and usually lower on Personal Maladaptation factors.

The Social and Emotional Domains of the CDER were reported to correlate with analogous domains of the Behavior Development Survey (Neuropsychiatric Institute Research Group, 1979), and in particular, the Emotional Domain correlated quite strongly (r=.78) with maladaptive factors on the Behavior Development Survey which is described elsewhere in the present report.

To summarize, the Emotional Domain comprises only a small part of the CDER. No published normative or standardization data could be found, and data were similarly missing with respect to internal consistency and test-retest reliability. Due to the method of reporting, it was difficult to assess the instrument's interrater reliability. Widaman et al.'s (1987) factor solution for the CDER suggests that it may be more appropriate to score the Emotional Domain items onto two subscales rather than one. There are some criterion validity data with the CDER, and these show a general tendency for the two maladaptation factors to correlate with the presence of a dual diagnosis. The instrument's congruent validity was difficult to assess, given the way that the data were summarized in the field manual. In conclusion, the Emotional Domain may have a role as a screening tool for behavioral/emotional disorders. However, it appears to be somewhat untested psychometrically, and it also seems less refined than many of the other instruments reviewed in this report. No doubt, this reflects the fact that the instrument was developed largely with other objectives in mind, namely, to provide data for management and administrative decisions.

References

Arias, M., Ito, E., & Takagi, N. (1983). *Concurrent validity of the Client Development and Evaluation Report.* Final report to the California Department of Developmental Services.

Department of Developmental Services (1986). *CDER Client Development Evaluation Report.* Sacramento, CA: Author.

Harris, C. W., Eyman, R. K., & Mayeda, T. A. (1982, June). *An inter-rater reliability study of the Client Development Evaluation Report.* Final report to the California Department of Developmental Services.

Neuropsychiatric Institute Research Group at Lanterman Hospital (1979). *Behavior Development Survey user's manual.* Unpublished manuscript available from Neuropsychiatric Institute Research Group, University of California at Los Angeles.

Nihira, K., Price-Williams, D. R., & White, J. F. (1988). Social competence and maladaptive behavior of people with dual diagnosis. *Journal of the Multihandicapped Person, 1*, 185-199.

Widaman, K. F., Gibbs, K. W., & Geary, D. C. (1987). Structure of adaptive behavior: I. Replication across fourteen samples of nonprofoundly mentally retarded people. *American Journal of Mental Deficiency, 91*, 348-360.

Clinical Interview Schedule
(also called "Standardized Psychiatric Interview")

D. P. Goldberg, B. Cooper, M. R. Eastwood, H. B. Kedward, &
M. Shepherd, 1970;
Modified for Persons with Mental Retardation by
B. R. Ballinger, J. Armstrong, P. J. Presley, & A. H. Reid,
1975

Point-form Synopsis

Stated Purpose: To measure abnormalities or changes in mental state (enabling an ICD-9 psychiatric diagnosis to be made, if relevant) in the context of survey-type interviews.

Age Range: Adults.

Levels of Mental Retardation Covered: Mild through profound.

Raters/Diagnosers: Experienced psychiatrists (Goldberg, Cooper, Eastwood, Kedward, & Shepherd, 1970).

Time Required to Complete: Thirty-five to 60 minutes (Fraser, Leudar, Gray, & Campbell, 1986).

Disorders/Dimensions Identified: Ten psychiatric symptoms elicited from the person by interview, and 12 manifest abnormalities observed during the interview (Goldberg et al., 1970). The modified schedule by Ballinger, Armstrong, Presley, & Reid (1975) contains 19 manifest abnormalities.

Date of Manual Publication: No manual for the modified interview, which is undated. Date of first publication with modified questionnaire: 1975.

Cost: Unknown. Modified questionnaire not commercially available.

Source: (Original Manual and Instrument): General Practice Research Unit, Institute of Psychiatry, London, England.
(Modified Questionnaire): Dr. Brian R. Ballinger, Consultant Psychiatrist, Royal Dundee Liff Hospital, Dundee DD2 5NF, Scotland, United Kingdom.

Limitations/Exclusions: Not researched on children with mental retardation. Some concepts asked not comprehended even by subjects having conversational skills; the concept of time in relation to symptoms poorly understood by many mentally retarded individuals (Ballinger et al., 1975). Interviewing may require engaging the person in appropriate play activities (Reid, Ballinger, Heather, & Melvin, 1984) and questioning caretakers about symptoms. Developers of the interview state that it must be administered by experienced psychiatrists with special training (Goldberg et al., 1970).

Description

The Clinical Interview Schedule (also referred to as the Standardized Psychiatric Interview) is a structured interview that was originally developed for use in community surveys with nonretarded populations (Goldberg et al. 1970). The interview is set out in four sections. Part 1 is relatively unstructured and contains questions concerning the person's present and past history regarding certain medical and psychiatric problems. Part 2 is a highly structured interview in which the interviewer asks the individual about 10 sets of symptoms. If the person responds affirmatively to a given item suggesting the presence of the symptom, there is a series of branching questions designed to elaborate on the details and seriousness of the symptom. Part 3 is relatively unstructured and contains additional questions about the individual's family and personal history. The interviewer has wide-ranging scope to explore any areas that may assist him or her in this part of the clinical assessment. Finally, Part 4 contains a list of abnormalities which may have been manifested during Parts 1 to 3. Each of these is rated on a 5-point scale once the person has left the room.

In the original report of the Clinical Interview Schedule (CIS), Goldberg et al. (1970) made reference to a clinical manual that contains instructions for conducting the interview, guidance for using the 5-point rating scales, and detailed descriptions and definitions of each symptom assessed. Goldberg et al. stated that the CIS should be administered only by experienced psychiatrists with special training in its use. The interview is designed to provide the necessary information to enable the interviewer to make an ICD (International Classification of Diseases) psychiatric diagnosis. However, the schedule is said to be more sensitive to neurotic than psychotic symptoms (Leudar & Fraser, 1987). The developers of the CIS list several possible objectives for the schedule as follows: (1) Use in large-scale community surveys as the second step in case-finding procedures, (2) Application within a defined population sample to test for associations between psychiatric disturbance and other variables, (3) The measurement of change in psychiatric state over a given time interval, and (4) Assessment of different population samples for comparing symptomatology and/or prevalence.

Ballinger and his associates (1975) modified the CIS for assessing adults with mental retardation. For purposes of this review, Part 2 (Symptoms Reported by the Subject) and Part 4 (Abnormalities Manifested During the Interview) are the critical domains, as they are used for classification and reporting purposes. Part 2 of the modified schedule contains the same 10 symptoms addressed in the original interview, as follows: (1) Somatic symptoms, (2) Fatigue, (3) Sleep disturbance, (4) Irritability, (5) Lack of concentration, (6) Depression, (7) Anxiety and worry, (8) Phobia, (9) Obsessions and compulsions, and (10) Depersonalization. Part 4 of the modified interview, relating to abnormalities manifested during the interview, contains the following 19 symptoms: (1) Slow, lacking spontaneity, (2) Suspicious, defensive, (3) Histrionic, (4) Depressed, (5) Anxious, agitated, tense, (6) Elated, euphoric, (7) Flattened, incongruous, (8) Delusions, thought disorders, misinterpretations, (9) Hallucinations, (10) Intellectual impairment, (11) Excessive concern with bodily functions, (12) Depressive thoughts, (13) Overactivity, (14) Distractibility, (15) Stereotypies, (16) Hostile irritability, (17) Lability of mood, (18) Pica, and (19) Self-injury. The first 12 of these manifest abnormalities are identical to those on the original interview (Goldberg et al., 1970), whereas the last seven were added by Ballinger and his associates. All symptoms are rated on a scale ranging from 0 through 4. As described by Ballinger and Reid (1977) and in the modified instrument, the scale is used as follows: (0) indicates absence of a symptom or manifest abnormality; (1) signifies a habitual trait or borderline symptom that does not cause significant distress or require treatment; (2) means that symptom is present in degree just sufficient to be considered pathological; (3) is recorded if the symptom is present in extreme degree intermittently or to a pathological degree persistently; (4) refers to extreme and persistent symptoms. Following Part 4, the interviewer is required to perform an Overall Severity Rating (using the same 5-point scale) and to formulate an ICD diagnosis based on the full interview. It is clear from discussions of the CIS that the interviewer is presumed to be a psychiatrist. However, at least one report included a clinical psychologist as a rater, and she obtained reliability levels as high as two psychiatrists who also were employed in the study (Ballinger et al., 1975).

Critique

The psychometric characteristics for the modified schedule are presented in Table 2 and Appendix B. The samples studied have covered the range of mental retardation from mild through profound. Average symptom scores and standard deviations are available on 133 subjects having mental retardation ranging from mild to severe (Fraser et al, 1986). This should be helpful in formulating clinical and research decisions about individual subjects.

To the writer's knowledge, there are no data relating to the internal consistency of this instrument. Test-retest reliability data are lacking, although one investigation reported 6-year follow-up data on a large group of institutional residents (Reid et al., 1984). In this study, 11 of 13 manifest abnormalities were correlated significantly over the 6 years, with five of the correlations (tau B) equal to or exceeding .50. Interrater reliability has been addressed in two studies involving samples of mentally retarded persons. In the first study (Ballinger et al., 1975), 27 subjects were rated by each of three raters. Correlations (derived from Analysis of Variance tables) ranged from -.18 to .93 (mean .64) for Part 2 symptoms. Correlations for Part 4 abnormalities ranged from -.02 to .69 (mean .20). Although Ballinger et al. (1975) regarded 20 of the 31 items to be satisfactory or very satisfactory, 10 of 12 symptoms and only 2 of 17 manifest abnormalities (12%) achieved reliability levels greater than or equal to .50 on Parts 2 and 4, respectively. In the second report, interrater reliabilities were reported for a small group of subjects (Fraser et al., 1986). The reliability for all of Part 2 (symptoms) was .78, and for Part 4 (abnormalities) it was .85.

In terms of its factorial/taxonomic validity, the suitability of the CIS hinges in large part on the ability of its items to derive an accurate ICD diagnosis. There are no instructions in the modified questionnaire as to how ratings should be translated into such a diagnosis. As is the case with the DSM-III, the appropriateness of the ICD psychiatric classification to the full range of mental retardation is unknown at this time. CIS data also have been used to produce a cluster solution (Reid, Ballinger, & Heather, 1978) and a factor solution (Fraser et al., 1986). Given time, comparison of these solutions with other empirically derived solutions may provide further support for the factorial validity of the interview, although their relevance to the structure of abnormal behavior in this population is presently unknown. A modicum of criterion group validity information comes from a comparison of subjects residing in institutions with other subjects living in the community. Those residing in institutions had significantly higher rates of a variety of acting-out symptoms (Ballinger & Reid, 1977). There is also a modest amount of congruent validity data with the CIS. In one investigation, a moderate level of agreement (r=.55) was found on Overall Severity ratings between researchers using the interview and consultant psychiatrists who had worked previously with the patients under study (Ballinger et al., 1975). In another report (Fraser et al., 1986), certain factors derived from the CIS were found to correlate with ratings on the Behavior Disturbance Scale (Leudar, Fraser, & Jeeves, 1984), but the two methods of obtaining information were weakly related overall.

Finally, a word is in order regarding the phraseology of the schedule and the ability of persons having mental retardation to respond to such items. Part 2, relating to reported symptoms, calls for the elicitation from the patient of any psychiatric symptoms he or she may have experienced in the preceding week. Ballinger et al. (1975) noted that some of the concepts in these questions seldom were grasped by their subjects, especially items regarding obsessions, compulsions, and depersonalization. Likewise, the concept of time (i.e., whether the symptom was present "within the last week") rarely was understood by their subjects. Ballinger and his associates attempted to deal with these problems by making appropriate adjustments, such as through interviews with caretaking staff about symptoms and by engaging the subject in appropriate play activities. Nevertheless, it is difficult to see how accurate information could be derived from individuals having severe and profound mental retardation, especially any concerning symptoms involving thought content and introspection.

To conclude, there appear to be several problems in employing the Clinical Interview Schedule with persons having the full range of mental retardation. Thus far, internal consistency data are lacking. However, this really may not be a problem, as it can be argued that such an interview is designed to cover the complete range of possible psychiatric abnormalities in the shortest possible time. Thus, individual questions would not be expected necessarily to correlate with one another. Data on interrater reliability suggest that it is not satisfactory for certain specific symptoms, especially for the manifest abnormalities section. Generally, there is a lack of information on test-retest reliability, although one follow-up study provides some suggestive data. The instrument's validity hinges largely on its suitability for yielding accurate ICD diagnoses and, in turn, on the relevance of ICD psychiatric symptoms and classifications to all levels of mental retardation. Evidence of criterion group and congruent validity is relatively weak at this time. Furthermore, some of the wording and concepts are probably beyond the comprehension of many persons with mental retardation.

Advantages of the instrument include the fact that it has been modified for, and has been used to a fair degree with, samples having mental retardation. Items from the schedule (especially the original questionnaire developed by Goldberg et al., 1970) appear to be well defined, with helpful descriptions to minimize ambiguity. Average scores for each of the symptoms also are available to assist in interpreting individual findings. It appears that more research (and possibly refinement) is needed before this instrument can be endorsed for broad application in this field. In particular, the reliability of individual symptoms needs to be addressed as well as criterion group and congruent validity. It may well be found that the interview proves to be very useful among mildly retarded individuals but that its effectiveness breaks down with persons having severe and profound retardation.

References

Ballinger, B. R., Armstrong, J., Presley, A. J., & Reid, A. H. (1975). Use of a standardized psychiatric interview in mentally handicapped patients. *British Journal of Psychiatry, 127*, 540-544.

Ballinger, B. R., & Reid, A. H. (1977). Psychiatric disorder in an adult training centre and hospital for the mentally handicapped. *Psychological Medicine, 7*, 525-528.

Fraser, W. I., Leudar, I., Gray, J., & Campbell, I. (1986). Psychiatric and behaviour disturbance in mental handicap. *Journal of Mental Deficiency Research, 30*, 49-57.

Goldberg, D. P., Cooper, B., Eastwood, M. R., Kedward, H. B., & Shepherd, M. (1970). A standardized psychiatric interview for use in community surveys. *British Journal of Preventive and Social Medicine, 24*, 18-23.

Leudar, I., & Fraser, W. I. (1987). Behaviour disturbance and its assessment. In J. Hogg & N. V. Raynes (Eds.), *Assessment in mental handicap: A guide to assessment practices, tests, and checklists* (pp. 107-128). London: Croom Helm.

Leudar, I., Fraser, W. I., & Jeeves, M. A. (1984). Behaviour disturbance and mental handicap: Typology and longitudinal trends. *Psychological Medicine, 14*, 923-935.

Reid, A. H., Ballinger, B. R., & Heather, B. B. (1978). Behavioural syndromes identified by cluster analysis in a sample of 100 severely and profoundly retarded adults. *Psychological Medicine, 8*, 399-412.

Reid, A. H., Ballinger, B. R., Heather, B. B., & Melvin, S. J. (1984). The natural history of behavioural symptoms among severely and profoundly mentally retarded patients. *British Journal of Psychiatry, 145*, 289-293.

Devereux Adolescent Behavior Rating Scale

G. Spivack, P. E. Haimes, & J. Spotts, 1967

Point-form Synopsis

Stated Purpose: To provide a means by which informants having a thorough knowledge of the youngster concerned can reliably describe and communicate overt problem behaviors in the individual being rated.

Age Range: Thirteen to 18 years.

Level of Mental Retardation Covered: Not reported. Developmental sample included mentally retarded adolescents, but IQ ranges not listed.

Raters/Diagnosers: Responsible adults with a good knowledge of the adolescent, such as parents, work supervisors, nurses, hospital aides, houseparents, and so forth.

Time Required to Complete: Approximately 10 minutes.

Disorders/Dimensions Identified: Twelve "factors" and three "clusters" are scored. (A) Factors: (1) Unethical Behavior, (2) Defiant-Resistive, (3) Domineering-Sadistic, (4) Heterosexual Interest, (5) Hyperactive-Expansive, (6) Poor Emotional Control, (7) Need Approval, Dependency, (8) Emotional Distance, (9) Physical Inferiority-Timidity, (10) Schizoid Withdrawal, (11) Bizarre Speech and Cognition, and (12) Bizarre Action. (B) Clusters: (1) Inability to Delay, (2) Paranoid Thinking, and (3) Anxious Self-Blame.

Date of Manual Publication: 1967.

Cost: Manual is priced at $2.00, whereas the unit cost of rating forms varies with the number ordered: package of 25, $7.50; 50 units, $13.00; 200, $44.00; and 500, $100. Postage and shipping extra.

Source: The Devereux Foundation, 19 South Waterloo Road, Box 400, Devon, PA 19333. Telephone (215) 964-3000.

Limitations/Exclusions: Probably not appropriate to the full range of mental retardation, especially severe and profound mental retardation.

Description

The Devereux Adolescent Behavior (DAB) Rating Scale is an informant instrument for rating the behavior of youth aged 13 to 18 years. The instrument was derived empirically by factor and correlational analyses from ratings on a mixed sample of 640 adolescents residing in several institutions. The samples included "disturbed" youths (many of whom had IQs well below normal), mentally retarded individuals (N=140), and normal adolescents . All individuals were rated on a form comprising 172 items. Items had been gleaned from the clinical literature, examination of child and adult rating scales, interviews with caregiving staff, and from clinical records. Due to computer limitations at the time, only 125 items could be included in the factor analysis. This resulted in an 18-factor solution. The additional 47 items, which had been held out of the analysis on the basis of very high or low correlations with the remainder of the item pool, were analyzed into four rational "clusters" on the basis of close intercorrelations between themselves and independence from the computer factors. Thus, this pair of analyses suggested a total of 22 fairly independent behavioral dimensions.

The actual DAB Rating Scale comprises 12 factor scores (subscales) and three behavior clusters (also subscales) derived from the empirical analyses, as well as 11 items that were retained because of their possible clinical and research value. The DAB Rating Scale is made up of a total of 84 items. Its 12 behavior factors have been designated as follows: (1) Unethical Behavior (4 items), (2) Defiant-Resistive (4 items), (3) Domineering-Sadistic (4 items) , (4) Heterosexual Interest (6 items), (5) Hyperactive-Expansive (6 items), (6) Poor Emotional Control (5 items), (7) Need Approval, Dependency (4 items) , (8) Emotional Distance (4 items), (9) Physical Inferiority-Timidity (5 items), (10) Schizoid Withdrawal (4 items), (11) Bizarre Speech and Cognition (7 items), and (12) Bizarre Action (5 items). The three clusters include the following: (1) Inability to Delay (6 items), (2) Paranoid Thinking (4 items), and (3) Anxious Self-Blame (5 items). However, these factors and clusters are only loosely based on the empirical analyses. It is noteworthy that 7 of the 15 subscales (47%) have only four items, a point that will be addressed subsequently.

The instructions for the DAB Rating Scale call for the rater to compare the adolescent subject with normal adolescents of the same age and sex and to rate the behavior patterns for the previous two weeks. Raters can be any responsible adult with a good knowledge of the individual, including parents, work supervisors, nurses, hospital aides, house-parents, etc. The authors specifically suggest that it is not desirable for teachers and client therapists to conduct ratings with the DAB Rating Scale. All items are rated either on 5- or 8-point Likert scales ranging from *Never* (or *Not at all*) through to *Very frequently* (or *Extremely*). Higher scores signify more serious behavioral/emotional problems. The manual states that it takes about 10 minutes to fill in the instrument. The DAB Rating Scale is a close relative of the Devereux Child Behavior Rating Scale, which also is reviewed in this report. Both were among the earliest empirically derived rating scales in the mental retardation field.

Critique

The psychometric characteristics of the DAB Rating Scale are presented in Table 2 and Appendix B. Like the Devereux Child Behavior Rating Scale, raw score totals for each subscale are plotted onto a profile, which shows the means and standard deviation units for normal and clinical samples. Clinically, this is a very useful feature. Unfortunately, the manual and related publication (Spivack & Spotts, 1967) present remarkably few details about the characteristics of the developmental samples, particularly insofar as IQ levels are concerned. Hence, the relevance of the instrument to mentally retarded youths is uncertain. It appears to the reviewer that many of the items would not be appropriate for individuals with severe or profound mental retardation.

In terms of the scale's internal consistency, "factor reliability" was found to range from .57 to .86, with a mean of .77 (Spivack & Spotts, 1967). This would suggest adequate, although certainly not extremely high, internal consistency. Test-retest correlations over a 7-to-10 day period ranged from .53 to .91 across subscales, with a mean of .81 (Spivack, Haimes, & Spotts, 1967), which can be regarded as adequate to very good. Interrater reliability was assessed with samples of disturbed adolescents and normal subjects, and mean correlations across the scales of .40 and .43, respectively, were obtained (Spivack et al., 1967). The authors presented more favorable results with another statistic called the coefficient of agreement, but it seems that a standard measure should be employed for comparisons across instruments and studies encompassed within the present report. The correlations cited are cause for concern insofar as the instrument's interrater reliability is concerned.

As noted, the structure of the DAB Rating Scale is based on a factor analysis and correlational analysis of its items. It seems to the reviewer that both the initial 18-factor plus four-cluster solution (Spivack & Spotts, 1967) and the derivative 12-factor and three-cluster solution (Spivack et al., 1967) are probably too fine-grained and elaborate to be stable across multiple clinical samples (see also critique of Devereux Child Behavior Rating Scale, this report). Furthermore, the composition of the definitive scale is based only loosely upon the original empirical analysis, with substantial changes in (1) the number of factors scored, (2) number of items loading on a given factor, and (3) allocation of subscales to a given "factor" or "cluster" (i.e., cluster 3 was originally a factor). Finally, such a fine-grained approach has the result of producing subscales with only a few items each (e.g., 47% of subscales had four items) and, as scale length is often directly related to reliability levels (Nunnally, 1967), this may have contributed to the low interrater reliabilities observed. Finally, there is a modicum of criterion group validity for the DAB Rating Scale. Mean subscale scores are presented in the manual comparing various diagnostic subgroups from the disturbed sample with normal adolescents living at home. All except two subscales (both belonging to the "clusters") differentiated at least some of the clinical groups from the normals. However, the size of the differences often appeared to be clinically nonsignificant. No data could be found on congruent validity for the DAB Rating Scale.

In summary, the format used in reference to normal and clinical groups is a useful feature of the score sheet for the instrument. The sample sizes, especially for the mentally retarded group, appear to be quite small and may not permit a meaningful breakdown, such as by age and sex. The absence of IQ data makes it difficult to judge the relevance of the DAB Rating Scale to adolescents having various degrees of mental retardation. Internal consistency appears satisfactory overall, and test-retest reliability looks adequate to good, but interrater reliability appears to be relatively low. Its factorial validity appears to be open to question, given the complexity of the factor solution adopted. There are some data on the instrument's criterion group validity, but no congruent validity data could be located. The DAB Rating Scale has been among the most frequently used published rating scales in the past (Hufano, 1985), and there is an extensive body of literature relating to the scale, but primarily with nonretarded subjects (see Institute of Clinical Training and Research, 1989). This is understandable in light of the fact that it was among the earliest empirically derived behavior rating scales in this field, and it occupies an important historical place because of this. However, it does not seem to be as technically sound as its close relation (the Devereux Child Behavior Rating Scale), and more research is needed with respect to interrater reliability and validity in general before the instrument can be recommended for broad clinical or research application with mentally retarded youth. The Devereux Adolescent Behavior Rating Scale is currently being revised and restandardized, and the new scale is expected to be available in the Spring of 1992 (P. LeBuffe, personal communication, September 12, 1989). It is quite possible that the revised instrument will resolve a number of the questions raised above.

References

Hufano, L. D. (1985). Devereux Adolescent Behavior Rating Scale. In D. J. Keyser & R. C. Sweetland (Eds.), *Test critiques* (Vol. III, pp. 221-225). Kansas City, MO: Test Corporation of America.

Institute of Clinical Training and Research (1989). *The Devereux Behavior Rating Scales revision project: Bibliography on the Devereux Adolescent Behavior Rating Scale.* Unpublished manuscript, The Devereux Foundation, Devon, PA.

Nunnally, J. C. (1967). *Psychometric theory.* New York: McGraw-Hill.

Spivack, G., Haimes, P. E., & Spotts, J. (1967). *Devereux Adolescent Behavior Rating Scale manual.* Devon, PA: The Devereux Foundation.

Spivack, G., & Spotts, J. (1967). Adolescent symptomatology. *American Journal of Mental Deficiency,* 72, 74-95.

Devereux Child Behavior Rating Scale
G. Spivack & J. Spotts, 1966

Point-form Synopsis

Stated Purpose: To provide a means by which informants having a thorough knowledge of the child concerned can reliably describe and communicate overt symptomatic problem behaviors of the child.

Age Range: Six to 12 years.

Level of Mental Retardation Covered: Not reported in manual. Subjects of normal IQ and all degrees of mental retardation included in developmental sample.

Raters/Diagnosers: Individuals having a thorough knowledge of the child over a period of time (e.g., parents, house-parents, nurses, child care workers, and so forth).

Time Required to Complete: Approximately 10 to 20 minutes.

Disorders/Dimensions Identified: Seventeen subscales are derived, as follows: (1) Distractibility, (2) Poor Self-Care, (3) Pathological Use of Senses, (4) Emotional Detachment, (5) Social Isolation, (6) Poor Coordination, (7) Incontinence, (8) Messiness, (9) Inadequate Need for Independence, (10) Unresponsiveness to Stimulation, (11) Proneness to Emotional Upset, (12) Need for Adult Contact, (13) Anxious-Fearful Ideation, (14) Impulse Ideation, (15) Inability to Delay, (16) Social Aggression, and (17) Unethical Behavior.

Date of Manual Publication: 1966.

Cost: The manual is priced at $2.00, whereas the unit cost of rating forms varies with the number ordered: Package of 25, $7.50; 50 units, $13.00; 200, $44.00; and 500, $100.00. Postage and shipping extra.

Source: The Devereux Foundation, 19 South Waterloo Road, Box 400, Devon, PA 19333. Telephone (215) 964-3000.

Limitations/Exclusions: None identified.

Description

The Devereux Child Behavior (DCB) Rating Scale was one of the earliest (if not the earliest) empirically derived instruments for rating the behavior of mentally retarded individuals. The scale was developed in two major stages. In the first, Spivack and Levine (1964) compiled a pool of potentially useful items for describing problem behavior in children. This pool was pared from 850 to 68 items that then were used to rate 140 children. Factor analysis of these ratings resulted in 15 interpretable factors. In a follow-up to this, Spivack and Spotts (1965) increased the scale to 121 items and obtained ratings on 252 "atypical" children who resided in four residential institutions. The outcome of the study was a 20-factor solution and, together with the earlier investigation, a 17-subscale instrument ultimately was compiled that encompassed 97 items.

The DCB Rating Scale was designed to be completed by any adult who has a good knowledge of the child over a period of time. Raters may include parents, house-parents, nurses, child care workers, and so forth. Instructions for the scale call for the rater to consider the child's behavior over the last two weeks and, in doing so, to compare the child with normal children of his or her age. Frequency-based Likert scales are used to score all items [e.g., (1) *Never* to (5)*Very frequently*]. The type of scale varies across parts of the instrument, and 5-, 8-, and 9- point scales are used.

The first 10 subscales of the Child Behavior Rating Scale have been characterized as "behavior competence" subscales, whereas the last seven have been labeled as "behavior control" subscales. The various subscales have been designated as follows: (1) Distractibility (4 items), (2) Poor Self-Care (2 items), (3) Pathological Use of Senses (3 items), (4) Emotional Detachment (6 items), (5) Social Isolation (3 items) , (6) Poor Coordination and Body Tonus (4 items), (7) Incontinence (3 items), (8) Messiness, Sloppiness (3 items), (9) Inadequate Need for Independence (4 items), (10) Unresponsiveness to Stimulation (4 items), (11) Proneness to Emotional Upset (8 items), (12) Need for Adult Contact (5 items), (13) Anxious-Fearful Ideation (7 items), (14) "Impulse" Ideation (5 items), (15) Inability to Delay (6 items), (16) Social Aggression (4 items), and (17) Unethical Behavior (4 items). Seventy-five of the 97 DCB Rating Scale items actually are used for scoring the 17 subscales. The remaining 22 items (#70, 71, 78-97) were retained because the authors felt that they might provide additional detail that could be useful clinically or for research purposes. Higher scores generally signify more serious problem behavior. However, extreme scores in either direction on the Need for Adult Contact subscale may indicate a difficulty.

Critique

The psychometric characteristics for the DCB Rating Scale are summarized in Table 2 and Appendix B. This instrument is one of the few known to the reviewer that uses several types of Likert scales (5-, 8-, and 9-point) to assess severity and, furthermore, the direction of scoring (i.e, high numbers may signify either high or low frequencies) varies across parts of the instrument. The reviewer knows from personal experience that some raters find this confusing. The test booklet provides for conversion of raw total scores to standard score units, which in turn are scaled with respect to a normal sample and a clinical sample having behavioral and emotional problems. This visual referencing of a given child's scores to normal and abnormal samples is very helpful clinically. The manual presents average subscale scores for a sample of 252 disturbed children, 100 mentally retarded children, and 348 public school children presumed to be normal. Given that the age range for these samples was either 5 to 12 or 5 to 13 years depending upon the sample, these reference groups are probably too small. The manual does not provide average subscale scores, broken down by age or sex, for the reference groups.

In terms of the DCB Rating Scale's reliability, no data are provided on its internal consistency. Test-retest reliability data were reported for 1-week, 1-month, and 6-month intervals, with mean correlations across the subscales being obtained of .90, .85, and .60, respectively (Spivack & Spotts, 1966). These are very high, although basic procedural data such as how many children were rated, types of raters, etc., were not reported. Intraclass correlation coefficients compared the consistency between supervisor and house-parent ratings (Spivack & Levine, 1964). Correlations ranged from .77 to .93 across subscales (mean = .84), which is very high. However, this comparison was conducted with a predecessor of the DCB Rating Scale, which differed in some substantial ways from the definitive instrument.

The instrument's breakdown into 17 factors is based on two factor analytic studies, described earlier. It is not clear from the manual how the two factor solutions, which resulted in 15 and 20 factors, were resolved into the definitive 17 subscale instrument. More importantly, however, this breakdown of the rating scale into 17 subscales may be overly fine-grained and, therefore, its structure may not be a robust behavioral representation for other clinical populations. For example, the present breakdown calls for separate subscales designated as Distractibility, Inability to Delay, and Poor Coordination, all dimensions that have been implicated as components of childhood hyperactivity. It is possible that a simpler solution may be more consistent with current knowledge of behavior disorders in children. Furthermore, the 17 factor solution results in 11 subscales having four or fewer items, and such small subscale sizes can have the undesirable effect of undermining reliability (Nunnally, 1967). Thus, like a previous reviewer (Polite, 1985), the writer's greatest concern about this instrument relates to what it actually measures. Criterion group validity was addressed by comparing children with mental retardation and behavioral/emotional disorders with controls, and the large majority of subscales showed differences in the expected directions (Spivack & Spotts, 1966). No data were reported on the congruent validity of the DCB Rating Scale with mentally retarded children.

To recap the foregoing, the standardization samples for the DCB Rating Scale were fairly small. The instrument's scoring system, which is referenced against normal and clinical samples, is a useful feature. The reviewer knows of no internal consistency data on this tool. Both test-retest and interrater reliability appear to be very high, although problems identified with both comparisons may limit their relevance. The intricate factor solution adopted for the DCB Rating Scale may prove to be unstable. The scale's criterion group validity appears to be acceptable, but no congruent validity data could be located. The DCB Rating Scale was one of the earliest standardized scales in the mental retardation field, and there is a sizeable body of literature on it, although most of it involves nonretarded samples (Institute of Clinical Training and Research, 1989). Furthermore, the DCB Rating Scale was one of the first instruments in this field to be derived in an empirical fashion. As such, the instrument occupies an important historical niche in the field, and it is perhaps regrettable that it has not been adopted more extensively for clinical research in the past. However, there are also possible technical problems and a lack of certain psychometric data with the DCB Rating Scale, and several of the more recent instruments are now likely to supersede it. It is worth noting, however, that like its close relative, the Devereux Adolescent Behavior Rating Scale, the DCB Rating Scale is currently being revised and restandardized, and the new instrument is expected to be available in the spring of 1992 (P. LeBuffe, personal communication, September 12, 1989).

References

Institute of Clinical Training and Research (1989). *The Devereux Behavior Rating Scales revision project. Bibliography on: The Devereux Child Behavior Rating Scale.* Unpublished manuscript, The Devereux Foundation, Devon, PA.

Nunnally, J. C. (1967). *Psychometric theory.* New York: McGraw-Hill.

Polite, K. (1985). The Devereux Child Behavior Rating Scale. In D. J. Keyser & R. L. Sweetland (Eds.), *Test critiques* (Vol. 2, pp. 231-234). Kansas City, MO: Test Corporation of America.

Spivack, G., & Levine, M. (1964). The Devereux Child Behavior Rating Scales: A study of symptom behaviors in latency age atypical children. *American Journal of Mental Deficiency, 68*, 700-717.

Spivack, G., & Spotts, J. (1965). The Devereux Child Behavior Scale: Symptom behaviors in latency age children. *American Journal of Mental Deficiency, 69*, 839-853.

Spivack, G., & Spotts, J. (1966). *Devereux Child Behavior Rating Scale manual.* Devon, PA: The Devereux Foundation Press.

Diagnostic Assessment for the Severely Handicapped (DASH) Scale

J. L. Matson, W. I. Gardner, D. A. Coe, & R. Sovner, 1990a

Point-form Synopsis

Stated Purpose: To provide a comprehensive structured survey of the psychiatric problems of individuals with severe or profound mental retardation.

Age Range: Primarily adults and adolescents (J. L. Matson, personal communication, June 1990).

Raters/Diagnosers: Scale completed by mental health professionals who interview appropriate informants, such as relatives of the individual or direct-care staff members who know the individual well.

Time Required to Complete: Estimated by reviewer at 20-25 minutes. Rating time increases with the number of positive symptoms.

Disorders/Dimensions Identified: Thirteen subscales are included as follows: (1) Anxiety, (2) Mood Disorder-Depression, (3) Mood Disorder-Mania, (4) Pervasive Developmental Disorder/Autism, (5) Schizophrenia, (6) Stereotypies/Tics, (7) Self-Injurious Behaviors, (8) Elimination Disorders, (9) Eating Disorders, (10) Sleep Disorders, (11) Sexual Disorders, (12) Organic Syndromes, and (13) Impulse Control and Miscellaneous Behavior Problems.

Date of Manual Publication: 1990.

Cost: A cost of duplication fee is assessed.

Source: Dr. Johnny L. Matson, Department of Psychology, Audubon Hall, Louisiana State University, Baton Rouge, LA 70803-5501. Telephone (504) 388-4104.

Limitations/Exclusions: Developed solely for severely and profoundly retarded people. Available norms exist only for institutionalized populations. Generally not regarded as appropriate for children.

Description

Because of the newness of this instrument, the reviewer was hesitant about including it in Part I of this report. Despite its relative youth, however, there are more data on this tool than on many older ones, and therefore, a more detailed coverage seemed to be warranted.

The Diagnostic Assessment of the Severely Handicapped (DASH) Scale is a recently developed survey of psychological problems for assessing individuals with severe and profound mental retardation (Matson, Gardner, Coe, & Sovner, 1990a). The instrument is designed to be administered to third-party informants, such as direct care staff members, who have extensive contact with the individual to be rated, or a relative of the person being rated. The scale is intended to be administered by a mental health professional who typically interviews an appropriate informant, although the professional may complete the form if he or she knows the individual well.

The instrument is made up of two sections; namely, a portion related to background information (12 questions) and a behavior rating component made up of 96 items describing behavior problems or psychiatric symptoms. Items were derived from the DSM-III-R (American Psychiatric Association, 1987) and previously developed instruments such as the Aberrant Behavior Checklist (Aman, Singh, Stewart, & Field, 1985) and the Behaviour Disturbance Scale (Leudar, Fraser, & Jeeves, 1984). The items of the DASH are organized into 13 disorder groups, largely on the basis of the structure of the DSM-III-R, as follows: (1) Anxiety (8 items), (2) Mood Disorder–Depression (15 items), (3) Mood Disorder–Mania (7 items), (4) Pervasive Developmental Disorder/Autism (6 items), (5) Schizophrenia (7 items), (6) Stereotypies/Tics (7 items), (7) Self-Injurious Behaviors (5 items), (8) Elimination Disorders (2 items), (9) Eating Disorders (6 items), (10) Sleep Disorders (5 items), (11) Sexual Disorders (3 items), (12) Organic Syndromes (9 items), and (13) Impulse Control and Miscellaneous Behavior Problems (16 items). Items were selected on the basis of appropriateness for people with severe and profound mental retardation and understandability to relatively untrained informants.

Each behavioral item is scored separately on three dimensions; namely, frequency, duration, and severity. On the frequency dimension, the rater is asked to indicate how often each item has occurred in the last two weeks, using a scale scored (0) *not at all,*

(1) *between 1 and 10 times*, or (2) *more than 10 times*. On the duration dimension, each item is scored in terms of how long it has existed, and once again a 3-point scale is used: (0) *less than 1 month*, (1) *1 to 12 months*, or (2) *over 12 months*. Finally, severity is scored for the last two weeks only on a dimension with the following points: (0) *behavior has caused no disruptions or damage*, (1) *the behavior has caused no injuries or damage but it has interrupted the activities of others*, or (2) *the behavior has caused property damage or injury to the individual or another person*. The instructions state that, if the behavior is found to have a frequency of zero (0) over the last two weeks, then the remaining dimensions (duration and severity) are not rated. The writer estimates that it typically would take 20-25 minutes to fill in the DASH, but given this branching procedure (i.e., with duration and severity rated only if frequency is rated 1 or 2), completion time will naturally increase directly with the number of problems that are endorsed.

Critique

Psychometric data for the DASH, available at the time of this writing, are summarized in Table 2 and Appendix B. Thus far, data are available on 506 severely and profoundly retarded residents of state institutions who were rated on the DASH (Matson, Gardner, Coe, & Sovner, 1990b). Average subscale scores and standard deviation units are available for the three dimensions of frequency, duration, and severity. However, mean scores were presented for subscales rather than individual items which may be significant, because a diagnosis is sometimes based on responses to a single item. Internal consistency (alpha) ranged from .20 to .84 over subscales with a median value of .52. This, of course, suggests poor to mediocre consistency for some of the subscales and, as noted by the authors (Matson et al., 1989b), the lower internal consistency values suggest that items within some subscales probably do not tap unitary dimensions of behavior. No data were presented on test-retest reliability or for item total correlations. Interobserver agreement was assessed by having different interviewers question two different informants for each of 29 subjects. Using a percentage agreement statistic, rater agreement (across all items) was found to be .96 for severity, .95 for duration, and .91 for frequency. Although these appear to be high, it would be very desirable to have some more appropriate statistic to gauge reliability, such as the kappa coefficient, as percentage agreement takes no account of chance rates of occurrence.

At the time of this writing, diagnoses were established arbitrarily for the 13 subscales (Matson, Gardner, et al., 1990b). For five subscales (Anxiety, Depression, Mania, Autism/Pervasive Developmental Disorder, and Schizophrenia) a given diagnosis was assigned if more than half of the subscale items were rated as present. For the remaining eight subscales a diagnosis was assigned if *any* subscale item was rated as present. Using these criteria, 91% of the sample were diagnosed as exhibiting one or more disorders (Matson, Gardner, et al., 1990b). Standardization data are also provided for diagnoses based on these criteria (Matson, Gardner, et al., 1990b).

The taxonomic validity of the DASH rests upon its relationship to the DSM-III-R, from which its structure is largely derived. As noted in the Introduction to this review, there are serious difficulties in assuming consistency between the structure and presentation of mental disorders across the full range of mental retardation as compared with mental disorders as they exist in people of normal IQ.

In another paper, the items of the DASH were factor analyzed using ratings of the same 506 subjects discussed above (Matson, Coe, Gardner, & Sovner, 1990). The outcome was a 6-factor solution encompassing 41 items that accounted for 39% of the variance. The factors were labeled as (1) Emotional lability, (2) Aggression/Conduct disorder, (3) Language disorder/Verbal aggression, (4) Social withdrawal, (5) Eating disorder, and (6) Sleep disorder. Alpha coefficients for the six factors ranged from .62 to .80 (mean .69). Mean factor scores and standard deviations are presented for the 506 subjects, with the group partitioned by level of mental retardation and ambulatory status. At the time of this writing, no data were available with respect to the criterion group or congruent validity of the DASH.

In a different vein, this reviewer wonders about the appropriateness of the actual numeric scales within the DASH for rating some symptoms. To take a specific example, the very essence of stereotypic behavior is that it is performed repetitively, and a frequency scale that permits only two levels of gradation (*between 1 and 10 times* and *more than 10 times*) may not be sufficiently sensitive to subject differences. Similarly, on the severity dimension it is hard to see how stereotypy would achieve a severity of 2 (corresponding to property damage and/or injury). Nevertheless, a strong case can be made that persistent stereotypic behavior can be a very severe behavior problem in the sense that it may impede the individual's development and acceptance into society. In any event, these are empirical issues, and the question of adequacy of the rating dimensions is quite testable.

In summary, there are standardization data available for the DASH, although only for subscale averages at the time of this writing. Given that diagnoses can be assigned for the presence of single symptoms, a standardization based on individual items is desirable. Interrater agreement appears to be high, but this needs confirmation with an appropriate statistic. The internal consistency for individual subscales is of variable quality. Thus far, there appear to be no test-retest reliability data, or criterion group or congruent validity data on the DASH. A total of 91% of an institutionalized population was assigned one or more diagnoses with the instrument, which strikes this reviewer as high although there are no standards against which to compare such a figure. A factor analysis has been performed on the DASH, permitting users to employ an empirically-derived scoring scheme, and standardization data are available for the factor scores. The DASH is at a very early stage of development, and it may be premature to subject it to review so soon. Despite this fact, there is a surprising amount of data available on the scale; this is an instrument which holds a great deal of promise, provided that the appropriate psychometric studies are carried out.

References

Aman, M. G., Singh, N. N., Stewart, A. W., & Field, C. J. (1985). The Aberrant Behavior Checklist: A behavior rating scale for the assessment of treatment effects. *American Journal of Mental Deficiency, 89*, 485-491.

American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders* (3rd edition, revised). Washington, DC: Author

Leudar, I., Fraser, W. I., & Jeeves, M. A. (1984). Behavioural disturbance in mental handicap: Typology and longitudinal trends. *Psychological Medicine, 14*, 923-935.

Matson, J. L., Coe, D. A., Gardner, W. I., & Sovner, R. (1990). *A factor analytic study of the Diagnostic Assessment for the Severely Handicapped (DASH) scale.* Unpublished manuscript, Louisiana State University, Baton Rouge.

Matson, J. L., Gardner, W. I., Coe, D. A., & Sovner, R. (1990a). *Diagnostic Assessment for the Severely Handicapped (DASH) Scale (User manual).* Unpublished manuscript, Louisiana State University, Baton Rouge.

Matson, J. L. Gardner, W. I., Coe, D. A., & Sovner, R. (1990b). *A scale for evaluating emotional disorders in severely and profoundly mentally retarded persons: Development of the Diagnostic Assessment for the Severely Handicapped (DASH) scale.* Unpublished manuscript, Louisiana State University, Baton Rouge.

Emotional Disorders Rating Scale – Developmental Disabilities

C. Feinstein, Y. Kaminer, & R. Barrett,

1988

Point-form Synopsis

Stated Purpose: To evaluate disorders of mood and affect in developmentally delayed children and adolescents for aiding diagnosis and assessment of treatment.

Age Range: Not specified. Developed for children and adolescents.

Level of Mental Retardation Covered: Mild and moderate mental retardation.

Raters/Diagnosers: Child care workers with a good knowledge of the individual.

Time Required to Complete: Not reported. Estimated by reviewer at 8 to 12 minutes.

Disorders/Dimensions Identified: Eight subscales as follows: (1) Anxiety, (2) Hostility/Anger, (3) Psychomotor Retardation, (4) Depressive Mood, (5) Somatic/ Vegetative, (6) Sleep Disturbance, (7) Irritability, and (8) Elated/Manic Mood.

Date of Manual Publication: No manual available. The rating scale is dated 1988.

Cost: Not commercially available.

Source: Carl Feinstein, M. D., Emma Pendleton Bradley Hospital, 1011 Veterans Memorial Parkway, East Providence, RI 02915. Telephone (401) 434-3400.

Limitations/Exclusions: Not designed for adults or for children/adolescents with severe and profound mental retardation.

Description

The Emotional Disorders Rating Scale–Developmental Disabilities (EDRS–DD) is a 59-item informant rating instrument for assessing developmentally disabled children and adolescents with mild and moderate mental retardation (Barrett, personal communication, September 1989). The instrument was designed to assess emotion in its broad sense, including disorders of mood and acting-out problems. The scale has eight subscales as follows: (1) Anxiety (6 items), (2) Hostility/Anger (7 items), (3) Psychomotor Retardation (9 items), (4) Depressive Mood (14 items), (5) Somatic/Vegetative (3 items), (6) Sleep Disturbance (5 items), (7) Irritability (6 items), and (8) Elated/Manic Mood (9 items). The items for the Anxiety subscale and the various types of Affective Disorder were constructed to meet DSM-III criteria. The Hostility/Anger and Irritability subscales were based on the authors' clinical experience and were designed to assess dimensions of emotionality that frequently cause clinical problems in children and adolescents with developmental disabilities (Feinstein, Kaminer, Barrett, & Tylenda, 1988). The Depressive Mood subscale contains seven items that presume verbal ability on the individual's part and another seven that do not. The verbal items are applicable only to higher functioning persons. Another version of the instrument exists (the EDRS), which is for use with children and adolescents who do not have developmental disabilities.

Existing publications on the EDRS-DD do not report who the intended raters should be, although a subsequent communication indicated that child care workers are the intended raters (R. Barrett, personal communication, September 1989). The instructions for the EDRS-DD call for the rater to complete two ratings for each item: the first assesses the frequency of the behavior on a 4-point scale [(0) *never* to (4) *often*], and the second assesses the severity of the problem [(0) *no problem* to (3) *severe*]. According to the authors, the severity scale appears to be markedly more useful than the frequency ratings (R. Barrett, personal communication, February 1989). Subscale scores are calculated by totaling the individual items for each respective subscale, and higher scores signify more serious behavior problems. The authors describe the EDRS as useful for measuring state-related changes in affective behavior and for assessing treatment response (Kaminer, Feinstein, Seifer, Stevens, & Barrett, in press).

Critique

Available data relating to the psychometric characteristics of the EDRS–DD are summarized in Table 2 and Appendix B. The reviewer was unable to locate average subscale scores, standard deviation units, or percentiles for the scale. Data regarding the instrument's internal consistency are available from a study of children of normal IQ (Kaminer et al., in press). Coefficient alpha ranged from .00 to .86 (mean = .51), levels which must be regarded as generally low. Test-retest stability over one week ranged from

–.14 to .84 (mean = .39). Again, these levels appear to be low, although it must be noted that these ratings took place in the context of a therapeutic program, and it is possible that observable behavior actually changed markedly over this period. Interrater agreement was reported to range from 85% to 96% (for frequency ratings) and from 86% to 96% (for severity ratings) in one study (Feinstein et al., 1988). However, percentage agreement takes no account of chance levels of agreement (e.g., both raters may have used predominantly "0" ratings), which tends to inflate the apparent level of agreement. In the other study, interrater reliability was found to range from .62 to .82 across subscales (mean = .72), which are moderate-to-high levels.

The instrument's factorial/taxonomic validity rests largely on its relationship to the DSM-III, from which six of its eight subscales were derived. The possible problems inherent in applying diagnostic schemes developed on the normal population to the population of mentally retarded persons already has been discussed at length in the Introduction. Criterion group validity was demonstrated with children of *normal* IQ for the Non-Verbal Depression items and the Manic/Elated Mood subscale, both of which were significantly related to a clinical diagnosis of depression (Kaminer et al, in press). However, no data were reported for other subscales or other diagnoses. Finally, congruent validity was reported between the Depressed Mood-Verbal Items and ratings on the Hamilton Depression Rating Scale (Hamilton, 1960) and the Children's Depression Rating Scale (Poznanski, Cook, & Carroll, 1979). No congruent validity data were presented for the other seven subscales assessed in that study (Kaminer et al., in press).

In summary, relatively few psychometric data appear to be available on the EDRS–DD. No standardization data could be located, and the available data suggest that its internal consistency is not high. The interrater agreement for the EDRS–DD appears to be satisfactory, but the only data on test-retest reliability are not encouraging. Generally, information is lacking on the scale's validity. Taxonomic validity for the EDRS–DD hinges on the relevance of the DSM-III categories to the mentally retarded population. There are some criterion group and congruent validity data in relation to the Depression and Manic/Elated subscales, but analogous comparisons are lacking for most of the eight subscales. Thus, the psychometric characteristics of the EDRS--DD are largely unresearched at this time. Although this instrument appears to be convenient and has a balanced format in terms of the behaviors subsumed, the lack of pertinent data may discourage its use in research.

References

Feinstein, C., Kaminer, Y., & Barrett, R. (1988). *The Emotional Disorders Rating Scale: Developmental Disabilities*. Unpublished instrument, Emma Pendleton Bradley Hospital, East Providence, RI 02915.

Feinstein, C., Kaminer, Y., Barrett, R., & Tylenda, B. (1988). The assessment of mood and affect in developmentally disabled children and adolescents: The Emotional Disorders Rating Scale. *Research in Developmental Disabilities*, *9*, 109-121.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*, 56-62.

Kaminer, Y., Feinstein, C., Seifer, R., Stevens, L, & Barrett, R. P. (in press). An observationally based rating scale for affective symptomatology in child psychiatry. *Journal of Nervous and Mental Disease*.

Poznanski, E. O., Cook, S. C., & Carroll, B. J. (1979). A depression rating scale for children. *Pediatrics*, *64*, 442-450.

Minnesota Developmental Programming System (MDPS): Behavior Management Assessment

W. H. Bock & R. F. Weatherman, 1979

Point-form Synopsis

Stated Purpose: (1) To assess adaptive behavior in developmentally disabled people in order to facilitate development of individual habilitation plans, and (2) to describe behavior problems in order to assess needs for management and treatment.

Age Range: Not specified. Presumably all ages.

Level of Mental Retardation Covered: Not specified. Presumably all levels.

Raters/Diagnosers: Staff, unit, or consulting psychologists, unit directors, supervising behavior therapists, or staff members performing ratings under their supervision.

Time Required to Complete: Not reported. Estimated by writer at 8 to 20 minutes. Rating time increases with greater number of problem behaviors.

Disorders/Dimensions Identified: Twenty-four behavior categories (no subscales as such).

Date of Manual Publication: 1985, 1989 (varies with state).

Cost: Unknown. Depends in part on how many components of MDPS are employed.

Source: Bock Associates, Inc., Court International, Suite 312 North, 2550 University Avenue West, Saint Paul-Minneapolis, Minnesota 55114. Telephone (612) 645-5300.

Limitations/Exclusions: None identified.

Description

The Behavior Management Assessment (BMA) is a small part of a larger system called the Minnesota Developmental Programming System (MDPS), originally developed as an adaptive behavior assessment instrument. The MDPS includes (1) the Assessment of Behavioral Competence, (2) an inventory called the Medical Needs Assessment, and (3) the Behavior Management Assessment. Originally the Assessment of Behavioral Competence comprised 18 domains and a total of 360 items (20 items per domain). An alternative form was developed for very young and low functioning individuals. The Assessment of Behavioral Competence is very lengthy and requires 1 to 2 hours to complete (Bock & Weatherman, 1979). For this reason, a shortened version was developed, comprising eight domains of 10 items each. This instrument was originally called the Minnesota Developmental Programming System–Abbreviated Form (MDPS–AF), but more recently the label *Scales of Behavioral Development* has been adopted (Bock Associates, 1989; Olvera, Bock, & Silverstein, 1985). The Medical Needs Assessment is a 12-item inventory describing special requirements of the client, such as appliances needed, requirement for special diets, the use of injections, medications, and so forth. The MDPS has become a very widely used system, and it has been adopted in part or wholly by a number of states, including Illinois, Indiana, Louisiana, Massachusetts, Minnesota, New York, North Dakota, and Oregon (Warren Bock, personal communication, May 1989).

The Behavior Management Assessment (BMA) is a 24-item instrument that describes a variety of maladaptive behaviors and psychiatric symptoms (e.g., coercive sexual behavior, pica, verbal abuse, mania). This list was compiled by using feedback from a group of 27 behavioral psychologists who were asked to identify behavior problems occurring among developmentally disabled people who live in state-operated and community-based settings. The psychologists also provided consensus judgments on appropriate frequency rates and descriptions of relative severity levels of each item (Bock Associates, 1989; Bock, McGovern, Schalock, Blakeman, & Silverstein, 1985; Silverstein, Olvera, & Schalock, 1989). The manuals for the BMA describe each behavior problem or symptom in fairly concrete behavioral terms and, furthermore, several levels of severity also are described for most items. However, provision for different degrees of severity are not provided for 6 of the 24 symptoms; namely, mania, inappropriate affect, substance abuse, hallucinations, delusions, and stereotypical behavior. Raters are asked to complete the form *only for individuals whose behavior is sufficiently frequent or intense to require a behavior management program.* Only those items describing behaviors warranting behavior management are rated, the remainder being left blank. For each relevant item, the rater is asked to identify a frequency and a severity level (formatted in a row by column matrix) to describe the intensity of that behavior or symptom. For the six items described previously that specify just cne level of severity, only the frequency of the behavior can be rated. The outcome of each rating is twofold. First, scores from each behavioral area are summarized. Second, a Total Score is computed which can be used to compare the severity of behavior problems across clients.

Two principal purposes are described for the BMA (Silverstein et al., 1989). First, it is said to be useful diagnostically for identifying problems and for comparing the severity of behavior problems among clients. Second, when data are available for a complete facility, it may serve an administrative function by revealing facilities or facility areas in need of greater or less staffing.

Additional Features

As with all aspects of the MDPS, the Behavior Management Assessment is designed for computer scoring. (It is not clear, however, whether score sheets are available for hand scoring of the instrument.) Second, the computer output presents a listing of suggested behavior management procedures that have been shown to be effective for treating each respective problem behavior. Although not prescriptive, these may provide a useful framework for structuring an individual's treatment.

Critique

Although a great deal of psychometric data are available for the adaptive behavior aspects of the MDPS, the writer was able to locate relatively few data relating to the psychometric attributes of the Behavior Management Assessment. Owing to the widespread use of the System in several states, it would appear that data for the Assessment would be available for numerous subjects. However, average Assessment scores could be found for only relatively small sample sizes, and standard deviation units were not presented in the report concerned (Olvera et al., 1985). Thus, it is not clear what the empirical basis is for assigning intensity scores to given behavior problems and to the Total Score. It also is not clear what the relation of the Total Score is to a given subject's reference group.

The writer was unable to locate any data relating to the internal consistency or test-retest reliability of the Assessment. Silverstein et al. (1989) reported that the interrater reliability for the Total Intensity Scores was .66, which might be considered a moderately good level of agreement.

No data could be located in relation to the factorial or criterion group validity for the Assessment. Congruent validity was assessed in a highly unusual way. Correlations were calculated between Total Intensity Scores and the actual staff time invested in behavioral management in each of three states (Silverstein et al., 1989). Correlations ranged from .67 to .81, with a mean of .76, showing a good relation between Total Score and committed resources. However, no data were presented to show that Total Scores were related to other measures of psychopathology.

The principal strength of the Behavior Management Assessment is its adherence to concrete behavioral descriptions for each component behavior problem. However, there are some peculiarities about the composition of the scale as well. For example, the quantification of severity for some items seems arbitrary and appears to be unverified empirically. For some items, such as inappropriate affect, it is not clear why only one level of severity was adopted. The description of some problems, such as hyperactivity, does not seem consistent with the relevant research literature (for instance, there is no consideration of attentional problems for that symptom). Additionally, standardization and psychometric data either could not be located or are barely adequate.

The available materials on the Behavior Management Assessment suggest that administrative considerations (e.g., allocation of resources within a service delivery system) were a major impetus for its development, and it may be admirably suited to this purpose. Furthermore, most of the parent system, the MDPS, is directed to the assessment of adaptive behavior. Although not reviewed here, the psychometric aspects for these components appear to be well researched. Insofar as research applications of the Assessment are concerned, it would seem that its major function might be confined to screening purposes. However, it is difficult to envisage that the assessment would supersede other instruments specifically developed for that purpose (e.g., the Reiss Screen for Maladaptive Behavior). Likewise, little is known about the assessment's reliability and validity for indentifying *individual* behavior problems, and its research and clinical application in this regard appear to be open to question.

References

Bock Associates Inc. (1989). *LSBD assessment manual: Louisiana Scales of Behavioral Development.* St.Paul-Minneapolis: Author.

Bock, W.H., McGovern, J., Schalock, R., Blakeman, L., & Silverstein, B. (1985). *Client based rate setting: Funding the habilitation and care of the developmentally disabled.* Springfield, IL: Illinois Office of Health Finance.

Bock, W. & Weatherman, R. (1979). *The assessment of behavioral competence of developmentally disabled individuals: The MDPS.* Minneapolis: University of Minnesota.

Olvera, D.R., Bock, W.H., & Silverstein, B.J. (1985). *Defining client profiles for residential and day program alternatives for developmentally disabled persons residing in the State of Indiana.* St. Paul-Minneapolis: Bock Associates, Inc.

Silverstein, B.J., Olvera, D.R., & Schalock, R. (1989). *Allocating direct care resources for treatment of maladaptive behavior: The Staff Intensity Scale.* Unpublished manuscript available from Barton Silverstein, Ph.D., Illinois Institute for Developmental Disabilities, Chicago, IL.

Preschool Behavior Questionnaire
L. B. Behar & S. A. Stringfield, 1974a, 1974b

Point-form Synopsis

Stated Purpose: To screen children at an early age for symptoms or constellations of symptoms that suggest the emergence of emotional problems.

Age Range: Preschool children aged 3 to 6 years.

Level of Mental Retardation Covered: Developed for normal IQ children; two studies available with developmentally disabled children.

Raters/Diagnosers: Preschool teachers.

Time Required to Complete: Not reported. Estimated by reviewer at 4 to 6 minutes.

Disorders/Dimensions Identified: (1) Hostile-Aggressive, (2) Anxious-Fearful, and (3) Hyperactive-Distractible.

Date of Manual: 1974.

Cost: Manual, $4.00; 50 answer sheets and score sheets, $8.00; postage per package, $3.00.

Source: Dr. Lenore Behar, Department of Human Resources, Division of Mental Health, Mental Retardation & Substance Abuse Services, 325 N. Salisbury Street, Albemarle Bldg., Raleigh, NC 27611.

Limitations/Exclusions: Children not in preschool; not suitable for raters other than teachers.

Description

The Preschool Behavior Questionnaire (PBQ) was developed as a screening tool for rating preschool-age children in nursery schools, day care centers, and kindergartens. Twenty-six items from Rutter's (1967) teacher rating scale served as the basis of the PBQ, although 10 new items (describing problem behaviors that occur frequently in preschoolers but not in older children) were added before its development. The PBQ was standardized on a normal sample of 496 children and an emotionally disturbed sample of 102 children who attended specialized treatment centers. *Children who had mental retardation, autism, or other handicaps specifically were excluded from the standardization sample.* Six items that failed to distinguish between the normal and emotionally disturbed groups were deleted, leaving a 30-item scale. Factor analysis of the ratings of children in the standardization groups rendered three factors: (1) Hostile-Aggressive (10 items), (2) Anxious-Fearful (9 items), and (3) Hyperactive-Distractible (4 items). Seven items on the definitive scale are not included in any of the subscale totals, although they do contribute to a Total score.

The instructions for the PBQ ask teachers to rate the child in terms of whether the item *does not apply* (0), *applies sometimes* (1), or *frequently applies* (2). Behar (1977) warns that employing raters other than teachers may produce data that are difficult to interpret, as the scale was developed solely with teachers. Item scores are totaled to determine subscale scores and a Total score; higher scores signify more serious behavior problems. The psychometric data for the PBQ will not be tabulated in this review for studies using children of normal IQ. Suffice it to say that research with such youngsters has indicated satisfactory to very good reliability (both test-retest and interrater), factorial validity, criterion group validity, and (generally) congruent validity (Behar, 1977). However, there was some question as to whether PBQ scores corresponded adequately with direct observations of behavior (Behar, 1977).

Critique

In keeping with the focus of this review, studies are discussed only if mentally retarded/developmentally disabled individuals served as subjects. This excludes a substantial body of research for the PBQ. The reviewer could locate two studies employing the PBQ, which are summarized in Table 2 and Appendix B. Thus far, average subscale and Total scores are not available for mentally retarded children. No data on reliability could be located for the PBQ with developmentally disabled children.

Rheinscheld (1989) conducted a factor analysis of ratings on 203 developmentally delayed children, and the original factor structure reported by Behar and Stringfield (1974a) was largely validated. Twenty-one of 24 items (88%) continued to load on the same respective factors. Both studies (Hammer, Kimball, & Beck, 1989; Rheinscheld, 1989) produced evidence of congruent validity, but the focus in both instances was on Attention Deficit/Hyperactivity. Hammer et al. (1989) found that ratings on the Hyperactive-Distractible subscale correlated highly with (a) teacher ratings of DSM-III criteria for Attention Deficit Disorder with Hyperactivity and with (b) commission errors on a continuous performance task. Hyperactive-Distractible ratings, however, were not correlated with omission scores or with attentional measures derived during a playroom session. Rheinscheld (1989) found that teacher ratings of activity level on a Likert scale were correlated with Hyperactive-Distractible scores, but they also were correlated, equally strongly, with Hostile-Aggressive scores.

To conclude, there are substantial data on the PBQ with children of *normal* IQ, but psychometric data with developmentally disabled children are very limited thus far. Work with the latter population suggests that the originally derived factor structure is valid, and there are also some data concerning congruent validity of the Hyperactive-Distractible subscale. Obviously, much work needs to be done before the PBQ can be recommended for widespread research or clinical use with mentally retarded preschoolers. Furthermore, applications of the instrument appear to be somewhat narrow at this stage. It is described as a screening tool by its developers, and it measures only three discrete problem areas. On the other hand, it may be difficult to differentiate a larger number of problem clusters at this age. Nevertheless, the PBQ does appear to warrant more psychometric work in this population. It is one of very few preschool rating scales, and the data thus far are encouraging.

References

Behar, L. B. (1977). The Preschool Behavior Questionnaire. *Journal of Abnormal Child Psychology, 5*, 265-275.

Behar, L. B. & Stringfield, S. A. (1974a). A behavior rating scale for the preschool child. *Developmental Psychology, 10*, 601-610.

Behar, L. B. & Stringfield, S. A. (1974b). *Manual for the Preschool Behavior Questionnaire.* Durham, NC: Learning Institute of North Carolina.

Hammer, D., Kimball, T. L., & Beck, S. (1989). *An analysis of similarities between developmentally delayed and nondelayed preschool boys with attention deficit disorder in their differential responses to* *objective measures of vigilance and activity level.* Manuscript in preparation, Ohio State University, Columbus, OH.

Rheinscheld, T. L. (1989). *A factor analytic study of the Preschool Behavior Questionnaire with developmentally delayed children ages 3-6.* Unpublished master's thesis, Ohio State University, Columbus.

Rutter, M. (1967). A children's behaviour questionnaire for completion by teachers: Preliminary findings. *Journal of Child Psychology and Psychiatry, 8,* 1-11.

Prout-Strohmer Personality Inventory
H. T. Prout & D. C. Strohmer, 1989

Point-form Synopsis

Stated Purpose: This is a self-report instrument intended to identify maladaptive personality patterns.

Age Range: Adolescents and adults (14 years and older).

Level of Mental Retardation Covered: Mild mental retardation and borderline intelligence (i.e., Full Scale or Verbal IQ between 55 and 83 inclusive on standardized IQ test).

Raters/Diagnosers: Individuals with borderline IQ and mild mental retardation. Administration to be guided by paraprofessional under supervision of a professional.

Time Required to Complete: Thirty minutes.

Disorders/Dimensions Identified: Five clinical scales as follows: (1) Anxiety, (2) Depression, (3) Impulse Control, (4) Thought/Behavior Disorder, (5) Low Self Esteem. A Lie scale score also is calculated.

Date of Manual Publication: 1989.

Cost: Complete kit (manual, plus scoring templates and 25 test protocols and scoring booklets), $77.50; 25 test protocols, $17.00; 25 scoring booklets, $17.00; complete kit with computer scoring software, $202.50.

Source: Genium Publishing Corporation, Psychological Testing Division, Department PS9A, 1145 Catalyn Street, Schenectady, NY 12303-1836. Telephone (518) 377-8854; FAX (518) 377-1891.

Limitations/Exclusions: Adolescents and adults with moderate through profound mental retardation; children less than 14 years; extremely uncooperative or disturbed individuals whose behavior or emotional state obviates self ratings.

Description

The Prout-Strohmer Personality Inventory (PSPI) is a self-report instrument for adolescents, aged 14 years and older, and adults comfortable with spoken English. It was developed for persons having borderline IQs and mild mental retardation (i.e., IQs in the 55 to 83 range). According to the manual, the inventory can be completed validly by over 90% of such individuals.

The inventory is made up of 162 items that resolve onto five clinical scales as follows: (1) Anxiety (25 items), (2) Depression (36 items), (3) Impulse control (33 items), (4) Thought/Behavior Disorder (20 items), and (5) Low Self Esteem (20 items). The inventory also has a Lie scale (12 items) to assess the tendency of some people to present an overly favorable picture of themselves (i.e., to "fake good"). A procedure to check for response sets is provided as well. Many items appear to have been adopted or modified from the Piers Harris Children's Self Concept Scale (Piers & Harris, 1969). Items were written in such a way that responses indicating a personality problem are balanced across "yes" and "no" answers for each subscale. Higher subscale scores are indicative of more serious personality problems.

The instructions for the inventory call for it to be administered by persons holding a higher degree in the social sciences or education or by a paraprofessional working under such a person's supervision. Each item is read aloud, while the individual follows along in his or her booklet and marks the appropriate response (yes or no). Subjects can be tested singly or in small groups. The manual states that interpretation of inventory profiles should be done only by professionals possessing at least a master's degree in the behavioral sciences or in education. The authors regard the PSPI as an important source of clinical information to be complemented by other data, such as behavior rating scales, observational data, interview techniques, and so forth. A companion instrument, the Strohmer-Prout Rating Scale, is reviewed elsewhere in this report.

Additional Features

Software is available (for IBM compatible computers) that will score the PSPI and provide a descriptive interpretation of potential problem areas.

Critique

The available data on the PSPI's psychometric characteristics are summarized in

Table 2 and Appendix B. In general, the manual is very detailed, with the provision of tables to convert raw scores to percentiles being a nice feature. In presenting data on the samples studied, considerable attention was given to relevant population characteristics, such as gender, age, racial composition, geographic distribution, use of medication, and so forth. Although this does not constitute standardization as such, the developers at least took cognizance of potentially important population characteristics which, unfortunately, has not been common in scale development in this field.

The internal consistency for the inventory appears to be good, with alpha coefficients ranging from .77 to .89 (mean .84) across subscales. Likewise, item-total correlations were moderately high, with an overall mean correlation of .40 between all items and their respective subscales. Test-retest reliability for the PSPI is excellent, with a correlation range of .65 to .89 (mean .81) across subscales in one study and .66 to .85 (mean .80) in another.

The data on validity are more problematic. The developers used a "rational/clinical" method to determine the selection of subscales and allocation of items to those subscales. However, the manual does not indicate what diagnostic or conceptual system guided that process. A confirmatory factor analysis was used to substantiate the assignment of items to subscales, but very few details are provided about the parameters employed in this analysis. Relevant summary data, such as individual factor loadings, were not reported. Furthermore, a stricter standard was applied to the factor analysis than to the rational/clinical assignment of items (i.e., items had to correlate .10 or more over and above the correlation with their original subscales for the empirical approach to force reassignment to a new scale). In fact, the subscales are fairly strongly intercorrelated (mean correlation = .64), suggesting that a smaller number of subscales may be more appropriate. This is one of the few instruments for which real content validity data are reported. Content validity was addressed by involving 15 professionals in contributing items to the item pool and, subsequently, by using 12 experienced workers to rate the items in terms of how strongly the items reflected their underlying clinical dimensions. Criterion group validity generally was very modest. Criterion groups expected to have more personality problems (e.g., those independently diagnosed as having emotional disorders or psychoses) generally scored higher than the remaining individuals. However, these differences did not appear to be statistically significant (relevant inferential comparisons were not reported) and, furthermore, the group differences generally did not approach levels that most workers would regard as clinically significant. This lack of discrimination may reflect some of the problems of source variance alluded to in the Introduction when acceptable levels of interrater reliability were discussed, but it would seem, nevertheless, to lessen the clinical utility of the inventory. The data on congruent validity were much better, especially when all data were derived from the individuals themselves (e.g., self ratings of anxiety and depression). However, the correspondence between caretaker ratings and subjects' self-ratings on analogous dimensions was typically weak.

To summarize, the manual for the inventory appears to be quite thorough, and substantial data are reported on the standardization sample. Internal consistency and test-retest reliability appear to be high. Insufficient data are provided to assess factorial/taxonomic validity of the scale. However, the presence of moderately high intercorrelations between subscales suggests that one or more factors may be common to several of these dimensions. The inventory is one of the few scales in which content validity was addressed in a serious way during its development. Criterion group validity data are relatively weak. The quality of the congruent validity data generally was determined by the source of the validational ratings: high for self ratings and low to moderate for informant ratings. This reflects two recurring problems with scale development alluded to in the Introduction. First, different raters have different perspectives on the individuals being assessed. Second, the task of validating a new instrument often is complicated by the very reason for its development, namely the lack of other suitable scales. Finally, the PSPI shares one weakness with all self-rating instruments in this area: It often is not usable or valid with extremely disturbed or uncooperative individuals. Of course, those are often the very persons that one wishes to assess. A great deal of effort has gone into the development and refinement of this scale, and it is one of the better self-rating instruments available in this field. However, more data attesting to its validity are needed before its place in research and clinical practice can be determined.

References

Piers, E.V. & Harris, D.B.V. (1969). *Manual for the Piers Harris Children's Self Concept Scale (The way I feel about myself).* Nashville, TN: Counselor Recordings and Tests.

Prout, H.T. & Strohmer, D.C. (1989). *Prout-Strohmer Personality Inventory manual.* Schenectady, NY: Genium Publishing Corporation.

The Psychopathology Instrument for Mentally Retarded Adults (PIMRA)

J. L. Matson, 1988

Point-form Synopsis

Stated Purpose: To help diagnose psychopathological conditions in people who are mentally retarded and to help plan mental health treatment and to assess treatment in such individuals.

Age Range: Adolescents or adults.

Levels of Mental Retardation Covered: All levels for informant ("Ratings-by-Others") version. Adolescents and adults with mild mental retardation and some adults with moderate mental retardation for the self-report version.

Raters/Diagnosers: Informant version: Caretakers in residential units, teachers, teacher's aides, work supervisors, family members, and mental health professionals. Self-report version: Individuals able to comprehend and respond to items on the instrument.

Time Required to Complete: Not reported. Estimated by writer at 6 to 12 minutes for informant version. Substantially longer for self-report version.

Disorders/Dimensions Identified: Eight subscale scores as follows: (1) Schizophrenia, (2) Affective Disorder, (3) Psychosexual Disorder, (4) Adjustment Disorder, (5) Anxiety Disorder, (6) Somatoform Disorder, (7) Personality Disorder, and (8) Inappropriate Adjustment. A Total Score also is calculated.

Date of Manual Publication: 1988.

Cost: Specimen set (one manual, self-report and informant questionnaires, and scoring form), $15.00.

Source: International Diagnostic Systems, Inc., 15127 South 73rd Avenue, Suite H-2, Orland Park, Illinois 60462. Telephone (312) 532-6337.

Limitations/Exclusions: Not appropriate for children (specific ages not indicated in manual).

Description

The Psychopathology Instrument for Mentally Retarded Adults is a checklist of abnormal behavior intended for use with people who are mentally retarded and who also may be mentally ill. According to the manual for the PIMRA, the intended uses of the instrument include the following: (1) planning psychological treatment, (2) evaluating the effects of mental health treatments, and (3) diagnosing psychopathological conditions in persons with mental retardation.

The PIMRA comprises 56 items that were based on major categories from the Diagnostic and Statistical Manual III (DSM-III) of the American Psychiatric Association (APA, 1980). The items were selected so that seven items contribute to each of the eight subscales as follows: (1) Schizophrenia, (2) Affective Disorder, (3) Psychosexual Disorder, (4) Adjustment Disorder, (5) Anxiety Disorder, (6) Somatoform Disorder, (7) Personality Disorder, and (8) Inappropriate Adjustment. In addition, a Total Score is calculated based on the sum of all 56 items. Each item is scored as either True or False for the person being rated.

Two versions of the PIMRA are available, an informant ("Ratings-by-Others") and a self-report version. The informant version is completed by people who know the individual well, such as parents, teachers, residential caregivers, work supervisors, or mental health professionals. The self-report version typically is read aloud to the mentally retarded individual who rates himself or herself. The self-report version is intended to be completed by adolescents and adults with mild mental retardation and some adults with moderate retardation, provided that they are able to understand and respond to the items on the instrument. No criteria are set out as to how this should be determined. On the informant version, all affirmative responses (yes or true) are indicative of psychopathology and are given a weight of 1. With the self-report version, yes and no responses are counterbalanced, and a given item is assigned a score of 1 (positive for psychopathology) for either type of response according to a scoring system that is provided in the manual.

Critique

The psychometric characteristics of the PIMRA are summarized in Table 2 and

Appendix B. Substantial data are available on one level, at least, with a minimum of 11 reports appearing in relation to this instrument. Concerning the populations studied thus far, persons with borderline through severe mental retardation have been assessed. With the exception of one study (Iverson & Fox, 1989), the author is not aware of studies assessing the PIMRA that have incorporated profoundly retarded individuals. In an early study, the internal consistency of the PIMRA was satisfactory with mean alpha coefficient levels of .85 and .83 reported, respectively, for the self-report and informant versions (Senatore, Matson, & Kazdin, 1985). However, alpha coefficients were calculated only for the total instrument, and internal consistency data were not derived separately for the eight individual subscales. Subsequent reports did calculate alpha coefficients for each subscale and found lower levels of internal consistency with mean alpha correlation coefficients of .64 and .66 on the self-report and informant versions, respectively (Aman, Watson, Singh, Turbott, & Wilsher, 1986; Watson, Aman, & Singh, 1988). Furthermore, 84% of the computed alpha coefficients in the latter reports fell below a level of .70, which might be regarded as signifying adequate levels of internal consistency (Reiss, 1988). Subsequently, mean alpha values of .32 and .41 were reported for the self-report version (Tymchuk, 1989) and the informant version (Sturmey & Ley, 1990).

Results for item-total comparisons have been mixed. These originally were calculated for the total instrument rather than individual subscales (Senatore et al., 1985). Subsequent comparisons have found item-subscale comparisons to range from low (Sturmey & Ley, 1990) to moderate (Tymchuk, 1989; Watson et al., 1988). Nevertheless, a few items have failed to correlate with Total Scale scores or with their respective subscale totals (Senatore et al., 1985; Aman et al., 1986; Sturmey & Ley, 1990; Watson et al., 1988), and these probably deserve further research scrutiny. In the hands of its developers, the PIMRA has produced mild to moderately high test-retest reliability levels (M=.56 and .76 for self-report and informant versions, respectively) (Senatore et al., 1985). However, another group found the levels of test-retest reliability to be generally low for the self-report version, ranging from -.15 to .56, with a mean of .31 (Watson et al., 1988). Two reports addressed interrater reliability. One study that compared self-report with informant ratings found correlations ranging from –.05 to .58 across subscales, with a mean correlation of .19 (Watson et al., 1988). This appears to challenge the reliability of either the self-report or the informant version, as they cannot both be "correct" and fail to correlate. In another study, two sets of informant ratings were obtained on 19 subjects (Iverson & Fox, 1989). Percentage agreement was said to range from 70% to 95% across subscales, with an overall mean agreement of 80%. Furthermore, 89% agreement was obtained regarding the occurrence or non-occurrence of significant psychopathology. However, percent agreement takes no account of rate of occurrence and makes no adjustment for agreement based solely on chance. Hence, these figures may be suggestive of higher reliability than would occur with, for example, the kappa coefficient (Fleiss, Spitzer, Endicott, & Cohen, 1972).

In terms of factorial/taxonomic validity, the items of the PIMRA were adapted from the DSM-III (APA, 1980). However, for reasons alluded to in the Introduction, it cannot be assumed with confidence that conditions appearing in the general (nonretarded) population necessarily occur unchanged across the range of mental retardation. Furthermore, even if we accept that such conditions do occur irrespective of level of mental retardation, we have no evidence thus far that they would be expressed symptomatically in the same way. Matson et al. (1984b) and Watson et al. (1988) conducted factor analyses of both versions of the PIMRA. Between two and four factors were found, depending upon the version analyzed and the particular study. The factor structures were remarkably similar across studies, but the obtained factors failed to confirm the scoring scheme for the instrument (although there was a fair degree of overlap with some subscales such as Anxiety Disorder). In commenting upon this in the PIMRA manual, Matson (1988) stated that "many disorders in the DSM-III were not themselves established empirically as valid diagnostic entities among nonretarded persons" (pg. 10). However, this does not lessen the difficulty insofar as the validity of the PIMRA is concerned. If in fact the empirical validity of the DSM-III is open to question, that only further undermines the structure of the PIMRA which is based directly upon it.

The principal evidence for criterion group validity comes from a study showing that subjects with diagnosed psychopathology had significantly higher Total Scores than subjects with no such documentation (Senatore et al., 1985). However, this in no way addresses the principal purpose of the PIMRA, which is stated in both the manuals for the PIMRA and for the Reiss Screen; namely, *to identify specific psychopathological conditions* in persons suspected of having both mental retardation and mental illness. Other evidence for criterion group validity comes from the demonstration that subjects receiving psychotropic medications had higher scores on certain subscales than unmedicated subjects.

Evidence for the congruent validity of PIMRA comes from a study showing a high correspondence between Total Scores and ratings on a predecessor of the Reiss Screen (Davidson, 1988). Most of the published validity work dealing with specific subscales of the PIMRA has dealt with the Affective Disorder subscale. For the self-report version, an association has been shown between the Affective Disorder scale and self ratings on the Beck but not on the Zung, Thematic Apperception Test, MMPI, or Hamilton depression scores. With the informant version, correspondence has been shown between the Affective Disorder and self ratings of depression on the Beck, Zung, Hamilton, and PIMRA (Affective Disorder) scales (Kazdin et al., 1983; Helsel & Matson, 1988). Ratings on the PIMRA have also been compared with ratings on the Aberrant Behavior Checklist (ABC) (Sturmey & Ley, 1990). In general, there was a tendency for PIMRA subscales to correlate significantly with apparently analogous subscales on the ABC.

According to the manual, there are no norms for comparing PIMRA scores for diagnostic purposes. A caveat within the manual urges that the results from the PIMRA "be considered in the context of a complete case evaluation" (pg. 2). Nevertheless, it would seem that professionals will be hampered in their interpretation of the PIMRA without any guidelines concerning average scores and deviation units for each subscale.

To summarize the foregoing, although there is a substantial amount of data on this instrument, the PIMRA appears either to be lacking or unresearched in certain respects. Only a modicum of data concerning the scale's interrater reliability appear to be available. In addition, validity data are very sparse regarding subscales other than the Affective Disorder subscale. As such, little is known about the validity of the instrument in establishing the presence of *specific disorders*. This is an apropos observation, because the manuals for both the Reiss Screen and the PIMRA state that this instrument should be considered as a follow-on to the Reiss Screen in order to establish the type of diagnosis when the presence of dual diagnosis is suspected. In a sense, these weaknesses appear to be more of a problem with the way that the PIMRA has been promoted and marketed than with the scale itself. *As the PIMRA undertakes to diagnose specific psychiatric conditions, it must meet higher standards than tools whose only function is to detect the presence of psychopathology*. Finally, there appears to be a rather weak correspondence between the self-report and the informant versions. This writer feels that the available psychometric data are more supportive of the informant that the self-report version, at least for most subscales describing acting-out forms of problem behavior. Finally, the absence of normative data impedes interpretation of individual profiles that emerge in the instrument.

To conclude, the PIMRA may be a promising screening instrument, but the available data do not support use of the PIMRA as the principal tool for establishing the presence of a specific psychiatric diagnosis. The PIMRA probably is useful as a structured questionnaire to provide a standard set of information that may prove helpful in assisting the diagnostic process. At this stage it may be best to regard the PIMRA as a helpful tool for probing for problem areas, but it needs much more research before it can be accepted as the central component for determining a specific diagnosis.

References

Aman, M. G., & Singh, N. N. (1986). *Aberrant Behavior Checklist manual.* East Aurora, NY: Slosson Educational Publications.

Aman, M. G., Watson, J. E., Singh, N. N., Turbott, S. H., & Wilsher, C. P. (1986). Psychometric and demographic characteristics of the Psychopathology Instrument for Mentally Retarded Adults. *Psychopharmacology Bulletin, 22*, 1072-1076.

American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed). Washington, DC: Author.

Davidson, M. (1988). *Psychometric characteristics of the Checklist of Emotional Problems with Mentally Retarded Adults (CHEMRA).* Unpublished doctoral dissertation. University of Illinois, Chicago.

Fleiss, J. L., Spitzer, R. L., Endicott, J., & Cohen, J. (1972). Quantification of agreement in multiple psychiatric diagnoses. *Archives of General Psychiatry, 26*, 169-171.

Helsel, W. J., & Matson, J. L. (1988). The relationship of depression to social skills and intellectual functioning in mentally retarded adults. *Journal of Mental Deficiency Research, 32*, 411-418.

Iverson, J. C., & Fox, R. A. (1989). Prevalence of psychopathology among mentally retarded adults. *Research in Developmental Disabilities, 10*, 77-83.

Kazdin, A. E., Matson, J. L., & Senatore, V. (1983). Assessment of depression in mentally retarded adults. *American Journal of Psychiatry, 140*, 1040-1043.

Matson, J. L. (1988). *The PIMRA manual.* Orland Park, IL: International Diagnostic Systems, Inc.

Matson, J. L., Kazdin, A. E., & Senatore, V. (1984a). Diagnosis and drug use in mentally retarded, emotionally disturbed adults. *Applied Research in Mental Retardation, 5*, 513-519.

Matson, J. L., Kazdin, A. E., & Senatore, V. (1984b). Psychometric properties of the Psychopathology Instrument for Mentally Retarded Adults. *Applied Research in Mental Retardation, 5*, 81-89.

Reiss, S. (1988). *Reiss Screen test manual.* Orland Park, IL: International Diagnostic Systems, Inc.

Senatore, V., Matson, J. L., & Kazdin, A. E. (1985). An inventory to assess psychopathology of mentally retarded adults. *American Journal of Psychiatry, 89*, 459-466.

Sturmey, P., & Ley, T. (1990). The Psychopathology Instrument for Mentally Retarded Adults: Internal consistencies and relationship to behaviour problems. *British Journal of Psychiatry, 156*, 428-430.

Tymchuk, A. J. (1989). *Symptoms of psychopathology in mothers with mental retardation.* Unpublished manuscript, University of California at Los Angeles.

Watson, J. E., Aman, M. G., & Singh, N. N. (1988). The Psychopathology Instrument for Mentally Retarded Adults: Psychometric characteristics, factor structure, and relationship to subject characteristics. *Research in Developmental Disabilities, 9*, 277-290.

Reiss Screen for Maladaptive Behavior
S. Reiss, 1988a; 1988b

Point-form Synopsis

Stated Purpose: To assess the likelihood that a mentally retarded adolescent or adult has a significant mental health problem.

Age Range: Greater than or equal to 12 years.

Level of Mental Retardation Covered: Mild through profound.

Raters/Diagnosers: Ratings from two or more caregivers are required, except for research purposes. Teachers, work supervisors, caregivers in residential units, teacher's aides, residential unit supervisors, mental health professionals, and so forth.

Time Required to Complete: About 20 minutes.

Disorders/Dimensions Identified: Eight subscale scores as follows: (1) Aggressive Behavior, (2) Psychosis, (3) Paranoia, (4) Depression (Behavioral Signs), (5) Depression (Physical Signs), (6) Dependent Personality Disorder, (7) Avoidant Disorder; and (8) Autism. In addition, a Total Score comprises the 26 items of the eight subscales, and six "special" maladaptive behaviors also are scored.

Date of Manual Publication: 1988.

Cost: Specimen set (one manual and rating form), $25.00 plus shipping/handling.

Source: International Diagnostic Systems Inc., 15127 South 73rd Avenue, Suite H-2, Orland Park, Illinois 60462. Telephone (312) 532-6337.

Limitations/Exclusions: Not appropriate for subjects less than 12 years of age. Requires two or more raters for clinical use.

Description

The Reiss Screen for Maladaptive Behavior is a screening instrument designed to identify persons with mental retardation who are likely to have a significant mental health problem. According to its developer, the instrument has several potential uses, including (1) screening for dual diagnosis in a variety of settings (state, provincial, metropolitan, community-based or developmental centers, and high schools), (2) providing structured information for intake evaluations at mental health and psychiatric facilities, (3) serving as a research tool in dual diagnosis research, and (4) providing instructional material for training workshops and seminars on dual diagnosis.

The Reiss Screen is made up of 38 items. Twenty-six items load onto one or more of seven subscales, as follows: (1) Aggressive Behavior, (2) Psychosis, (3) Paranoia, (4) Depression (Behavior Signs), (5) Depression (Physical Signs), (6) Dependent Personality Disorder, and (7) Avoidant Disorder. Each of these scales comprises five items, although some items load onto more than one scale. Each scale was derived by factor analysis from data on a diverse sample of 306 persons, most of whom were dually diagnosed, in six states and the province of Ontario. Subsequent to the factor analysis, an Autism Scale was added, and this comprises a further five items. In addition to the eight subscales, there are also six "special symptoms" that describe serious behavior problems. These special symptoms include the following: (1) Drug/Alcohol Abuse, (2) Self-Injury, (3) Stealing, (4) Overactivity, (5) Sexual Problem, and (6) Suicidal Tendencies. There also are two experimental items on the Screen (i.e., items 14 and 36, not scored), bringing the total to 38 items. Finally, a 26-item Total Score also is calculated. This is based on the sum of the items forming the original seven subscales derived from the factor analysis, and it may be construed as a rough measure of the severity of psychopathology in a given case.

Each item is scored on a 3-point scale ranging from (0) *no problem*, through (1) *a problem*, to (2) *a major problem*. In scoring each item, raters are asked to take both frequency and severity into account. Detailed instructions and examples are provided to clarify how the rating scale should be used. The instructions require that each person being rated be evaluated by two or more raters who know the individual well, among whom may be teachers, work supervisors, family members, or any professionals meeting this criterion. The manual provides cutoff scores for the Total Score, each of the eight subscales, and for each of the six special symptoms. (Fourteen possible scores and cutoffs are provided.)

Additional Features

The marketer of the Reiss Screen, International Diagnostic Systems, offers three services to help with scoring the instrument: (1) Scoring forms to guide calculations are provided; (2) IBM-compatible software is available for personal computers; and (3) A computerized scoring service is available in which completed forms can be scored, providing a printout for each individual rated and a summary for the whole group.

Critique

The psychometric characteristics of the Reiss Screen are presented in Table 2 and Appendix B. In general, its psychometric properties have been well researched and appear to be substantially better than average. The normative sample was somewhat small (N=258), whereas the validation samples totaled to a more acceptable figure (N=418). The instrument was developed entirely with samples of mentally retarded persons whose level of retardation ranged from mild to profound. Generally, there was an attempt to include samples that were characteristic of the national population of retarded persons in terms of age, sex, race, and functional impairment. Internal consistencies were adequate for most subscales, with alpha coefficients generally above .70, although the Depression (Physical Signs) Scale had lower levels of internal consistency. Interrater reliability was reported only for individual items (rather than subscale totals), although the levels reported were generally very acceptable (M=.54). Validity was established by factor analysis, criterion groups, and congruent measurements with other instruments. In general, the evidence for validity is good insofar as the instrument is used for the identification of *any* psychopathology. Relatively few data have been presented to establish the instrument's utility for establishing a *specific* diagnosis. It should be reiterated, however, that the principal purpose of the Reiss Screen is to establish whether or not there is a need for further diagnostic assessments.

The principal drawbacks of the instrument appear to be threefold. First, the standardization group on which normative data are based appears somewhat small (N=258). This has implications for the confidence with which cutoff scores can be accepted. Second, the choice of cutoff score levels appears to have been somewhat arbitrary for some subscales and special symptoms. For example, the Total Score cutoff was set at a value that was approximately midway between the scores for a no-diagnosis group and a dual diagnosis group. The Autism cutoff was set high relative to scores for an autistic subgroup because of concerns that the symptoms for this group may have diminished with age. Third, the specificity of the various subscales seems somewhat low, in that subscale scores for groups of dually diagnosed subjects with a particular disorder did not always differ appreciably from subscale scores of individuals having entirely different types of dual diagnoses. Although it may be argued that the Reiss Screen was not designed to yield specific diagnoses, it does in fact use a diagnostic format, and some users will almost certainly attempt to employ it in this manner. Nevertheless, in spite of these reservations, it must be concluded that the Reiss Screen is a relatively well researched screening instrument, and the available psychometric data, in general, suggest that it compares favorably with most other available instruments in this field.

References

Davidson, M. (1988). *Psychometric characteristics of the Checklist of Emotional Problems with Mentally Retarded Adults (CHEMRA).* Unpublished doctoral dissertation, University of Chicago, Chicago, Illinois.

Reiss, S. (1988a, September). *The development of a screening measure for psychopathology in people with mental retardation.* Paper presented at the joint National Institute of Mental Health--National Institutes of Health Workshop on "The Assessment of the Mental Health Aspects of Mental Retardation." Bethesda, Maryland.

Reiss, S. (1988b). *Test manual for the Reiss Screen for Maladaptive Behavior.* Orland Park, IL: International Diagnostic Systems.

Sacristan, J. R. (1987). *Cross cultural aspects on epidemiology in assessment of dual diagnosis.* Paper presented at the International Research Conference on the Mental Health Aspects of Mental Retardation. Evanston, Illinois.

Schedule of Handicaps, Behaviour, and Skills (HBS)–Revised (Formerly called Schedule of Children's Handicaps, Behaviour, and Skills

L. Wing, 1982

Point-form Synopsis

Stated purpose: To serve as a framework for eliciting clinical information to describe the person's level of functioning and present behavior for assessment and diagnostic purposes.

Age Range: Originally developed for children. The revised schedule has been extended since to include adults (Wing, 1980).

Level of mental retardation covered: All levels (L. Wing, personal communication, December 1989).

Raters/Diagnosers: Professionals who have received training in use of the instrument and who are familiar with mentally retarded and autistic children.

Time Required to Complete: Forty-five minutes to 2 1/2 hours (Wing & Gould, 1978).

Disorders/Dimensions Identified (Behavioural Abnormalities Component): Fifteen sections, each of which may have several parts. See text.

Date of Manual Publication: 1978; revised in 1982.

Cost: No charge.

Source: Dr. Lorna Wing, MRC Social Psychiatry Unit, Institute of Psychiatry, DeCrespigny Park, London SE5, England. Telephone 01-703-5411 (Ext. 3502).

Limitations/Exclusions: Distribution of schedule is restricted to people with experience in autism and mental retardation.

Description

The Schedule of Handicaps, Behaviour and Skills (HBS) is a semistructured interview that was developed for trained professionals who are very familiar with mentally retarded and "psychotic" individuals (Wing & Gould, 1978). Originally developed for the assessment of children, the HBS has since been extended for use with adults. Its purpose is to provide all information that is necessary to arrive at a diagnosis and to develop a prognosis, but the schedule was primarily developed as a research tool to investigate autism. According to Bernsen (1980), the HBS Schedule was developed for use with children having moderate through profound mental retardation as well as youngsters who are retarded in some, but not all, aspects of their development.The revised schedule may be applied to children or adults with mild through profound mental retardation (Wing, 1980).

The structure of the HBS Schedule is difficult to decipher because earlier accounts of the instrument (e.g., Wing & Gould, 1978) and the layout of the schedule as provided to the reviewer appear to differ. The writer has assumed that the latest format that he obtained is correct and the following description is based on that assumption, but readers should bear in mind that there may be some minor inaccuracies.

Descriptions of the schedule speak of a Developmental Skills component and a Behavioural Abnormalities component. The HBS Schedule contains 33 separate sections, each of which may contain several questions. In addition there are appendices with four sections which also describe psychiatric disorders or behavior problems.

The Developmental Skills sections relate to functional skills and can be used to determine level of adaptive behavior. From perusal of the schedule it is not possible to determine with certainty which sections belong to the Behavioural Abnormalities Component, but the following sections appear to be relevant: (1) Abnormalities of Speech or Sign Language, (2) Abnormal Imaginative Activities, (3) Eye Contact, (4) Social Responsiveness, (5) Social Play, (6) Social Interaction, (7) Abnormal Response to Sounds, (8) Abnormal Response to Visual Stimuli, (9) Abnormal Proximal Sensory Stimulation, (10) Abnormal Bodily Movements, (11) Routines and Resistance to Change, (12) Behaviour Problems with Limited or No Social Awareness, (13) Behaviour Problems with Social Awareness (14) Sleeping Problems, and (15) Initiative and Perseverence. To give the flavor of these sections, the "Routines and Resistance to Change" section (number 11) includes the following elements: (a) Dislike of change in the normal routine, (b) Routines invented by the person, (c) Food fads, (d) Clinging to objects, (e) Interest in special objects or parts of objects, and (f) Special fears. The four sections of the

Appendix are listed as follows: (1) Abnormal Postures and Movements, (2) Sexual Problems, (3) Psychiatric Problems, and (4) Legal Problems. The third section, Psychiatric Problems, inquires about 12 psychiatric disorders, such as depression, mania, obsessions, schizophrenia, personality disorders, and so forth. However, the specific criteria for determining the presence of these disorders are not spelled out as they are in the earlier sections.

All items on the HBS Schedule are scored with respect to the person's behavior over the previous month. For the Developmental Skills portion, the person is rated according to the developmental level he or she has reached at the time of the interview. Higher developmental stages are coded with higher scores for the given subsection. Items on the Behavioural Abnormalities component are scored in the same manner, and (in contrast to most instruments reviewed in this report), higher scores indicate *less* abnormal behavior by the individual. On many sections comprising the HBS Schedule, items are scored on scales ranging from (0) *markedly abnormal behavior*, described in concrete terms, through (3) *normal behavior*. In at least some reports (Bernsen, 1980; Wing & Gould, 1978) the ratings from subsections have been combined to give a 3-point rating for each section. For the Behavioural Abnormalities sections, the lowest rating, 1, indicated that the problem existed to a marked degree; the intermediate rating, 2, indicated that it existed to a moderate degree; and a rating of 3 indicated that the problem was minimal or absent.

The HBS Schedule is a semistructured interview, and the interviewer has wide-ranging scope to probe for accurate information regarding a given item. However, introductory questions are provided for the various sections to facilitate the interview process. Interview time can vary greatly depending in part on how articulate and reliable the informant is and also on the complexity of the behavior of the person concerned. In one study, total interview time ranged from 45 minutes to 2 1/2 hours (Wing & Gould, 1978). However, it should be noted that this was for the full schedule, and interview times for the Behavioural Abnormalities section, if given alone, necessarily would be less.

Although the manual states that the schedule is designed to assess functional level and present behavior, the schedule places a very heavy emphasis on questions related to childhood autism (e.g., imaginative activities, eye contact, social responsiveness, abnormal bodily movements, etc.). As noted, the schedule also contains sections related to a variety of behavior problems both with and without a social context, and an appendix to the schedule also includes the gamut of abnormal sexual and psychiatric conditions. However, these are not explicated in detail as are the symptoms related to autism and other developmental problems, and it appears that a principal objective for the instrument was to evaluate what Wing (1981) refers to as the "triad of social and language impairment." This triad refers to abnormalities of social interaction, verbal and nonverbal communication, and imaginative activities.

Critique

Only the Behaviour Abnormalities component will be reviewed here. The psychometric characteristics for this part are summarized in Table 2 and Appendix B. Thus far, the HBS Schedule has been used in research both with children (Bernsen, 1980; Wing & Gould, 1978) and with adults (Lund, 1985). However, there appear to be no standardization or normative data for the instrument with large samples of mentally retarded persons (see Wing, 1980, for discussion).

In terms of the instrument's reliability, the reviewer could locate no data on its internal consistency, item-total correlations for the various sections, or test-retest reliability. Interrater reliability for the HBS Schedule has been assessed in two ways (Wing & Gould, 1978). First, diagnoser accuracy was evaluated by having two clinicians assess the same group of 20 children using audiotapes (i.e., the second examiner listened to the interviews conducted by the first examiner). Complete agreement was achieved across all 20 subjects for all except one section; namely, Repetitive Symbolic Play. Interrater reliability also was assessed by having the clinicians conduct independent interviews with two informants; namely, the children's mothers on the one hand and professional caretakers on the other (e.g., teachers, nurses, child care workers, training center supervisors). For this exercise, three unique indices were employed to assess informant agreement: (1) Maximum Agreement (MA) referred to the percentage of children on whom both the parents and professional informants gave the same section ratings. (2) Agreement for Presence (AP) referred to the number of children for whom *both* types of informant described a symptom as present divided by the number of children for whom *either* informant described the symptom as present. (3) Finally, Agreement for Absence (AA) referred to the number of children for whom both informants regarded the symptom as absent divided by the number regarded by either informant as absent. In general, as these indices approached 1.00, the level of agreement was regarded as higher. Substantial agreement appears to have been achieved depending upon the index used (see Appendix B). However, this is an extremely unwieldy method of reporting agreement, as the number of sections meeting a given criterion changes according to the number of children correctly classified by both informants. Furthermore, percentage agreement of this form takes no account of chance levels of agreement (Fliess, Spitzer, Endicott, & Cohen, 1972) and, as such, may suggest higher levels of agreement than is, in fact, the case.

The reviewer could locate no data on the instrument's factorial/taxonomic validity or its congruent validity. Possible evidence for criterion group validity comes from two reports that compared "psychotic" (Wing, 1978) and socially impaired (Wing & Gould, 1979) children with sociable children having mental retardation. It was found that the aloof children differed from the sociable children on a variety of HBS Schedule sections, including those related to eye contact, presence of stereotypies, elaborate routines, symbolic play, echolalia, language comprehension, organic conditions, and delay of onset after birth. However, the classification of unsociable vs. sociable appears to have been based on data from the schedule itself, so that these associations seem to reflect diagnostic clusters appearing within the instrument, rather than evidence of validity with an external criterion.

To summarize, there appear to be relatively few data on the psychometric properties of the HBS Schedule, and available data are confined to children. There are some data on interrater agreement, but the unconventional statistics used fail to take chance agreement into account, and they do not allow for comparison with other instruments. The only data that could be located regarding the instrument's validity appeared to reflect on relationships between sections contained *within* the instrument rather than with external criteria. The rules governing the actual interpretation of scores from the schedule are not specified in the materials obtained by the reviewer. It is not clear from reading publications relating to the instrument how the various sections were derived. Furthermore, the basis for determining the presence of a number of disorders summarized in Appendix B to the schedule is not spelled out, and the use of these categories is presumably consistent with the diagnostic system (such as the ICD-9) from which they were derived. Thus, despite a history of use of this instrument in both England and Denmark, its psychometric characteristics remain largely unstudied. In a discussion of the HBS Schedule, Wing (1980) noted that it is not a "psychometric" instrument (meaning that raw scores are not to be used in a simplistic fashion), and she stressed that clinical experience and judgment are important prerequisites for deriving valid diagnoses with this tool. Given the available data, its value as a *general* diagnostic tool in the mental retardation field remains to be demonstrated. However, it is the reviewer's impression that the instrument is probably useful for assessing a narrower group of disorders, such as autism and what Wing (1981) describes as the triad of social interaction, communication, and imagination.

References

Bernsen, A.H. (1980). An interview technique in assessing retarded children. *Journal of Mental Deficiency Research, 24*, 167-179.

Fleiss, J.L., Spitzer, R.L., Endicott, J., & Cohen, J. (1972). Quantification of agreement in multiple psychiatric diagnoses. *Archives of General Psychiatry, 26*, 169-171.

Lund, J. (1985). The prevalence of psychiatric morbidity in mentally retarded adults. *Acta Psychiatrica Scandinavica, 72*, 563-570.

Wing, L. (1975). A study of language impairments in severely retarded children. In N. O'Conner (Ed.), *Language, cognitive deficits, and retardation* (pp. 87-112). London: Butterworth.

Wing, L. (1978). Social, behavioral, and cognitive characteristics: An epidemiological approach. In M. Rutter & E. Schopler (Eds.), *Autism: A reappraisal of concepts and treatment* (pp. 27-45). New York: Plenum Press.

Wing, L. (1980). The MRC Handicaps, Behavior & Skills (HBS) Schedule. *Acta Psychiatrica Scandinavica*, Suppl. 285, 241-247.

Wing, L. (1981). Language, social and cognitive impairments in autism and severe mental retardation. *Journal of Autism and Developmental Disorders, 11*, 31-44.

Wing, L. (1982). *Schedule of Handicaps, Behavior and Skills.* Unpublished manual available from MRC Social Psychiatry Unit, Institute of Psychiatry, DeCrespigny Park, London, England SE5 8AF.

Wing, L., & Gould, J. (1978). Systematic recording of behaviors and skills of retarded and psychotic children. *Journal of Autism and Childhood Schizophrenia, 8*, 79-97.

Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders, 9*, 11-29.

Self-Report Depression Questionnaire
W. M. Reynolds, 1989

Point-form Synopsis

Stated Purpose: To assess the depth of depressive symptomatology reported by individuals with mental retardation.

Age Range: Adolescents and adults.

Level of Mental Retardation Covered: Not reported. Psychometric characteristics studied with mildly through severely retarded subjects.

Raters/Diagnosers: Individuals with mental retardation and/or brain injury able to understand and respond to scale items. Administration guided by trained clinical interviewers.

Time Required to Complete: Estimated by reviewer at 12 to 20 minutes.

Disorders/Dimensions Identified: A pretest, used to assess the person's ability to complete the inventory, and a total Depression score.

Date of Administration Booklet: 1985, 1989. Manual in preparation.

Cost: Unknown.

Source: Psychological Assessment Resources, Inc., P. O. Box 998, Odessa, FL 33556. Telephone (813) 968-3003; FAX (813) 968-2598.

Limitations/Exclusions: Not suitable for children or persons unable to understand or respond to Questionnaire items.

Description

The Self-Report Depression Questionnaire (SRDQ) was intended to provide an index of the depth of depressive symptomatology in adolescents and adults with mental retardation. The instrument is divided into two major sections: (a) a two-part pretest and (b) the questionnaire, which assesses symptoms of major and minor depression.

The pretest is intended to determine whether the person is capable of responding reliably to the questionnaire and, specifically, whether he or she can differentiate between the response choices (*almost never*, *sometimes*, and *most of the time*). The pretest is made up of five practice items (Part I) and a further 15 questions (Part II) comprising the pretest itself. The pretest is made up of statements that are rarely, sometimes, or usually true of the vast majority of people. Thus, "I get dressed when I wake up" predictably would be answered *most of the time* by the large majority of the population, whereas, "It snows in the summer" correctly would be answered *almost never* by the brunt of respondents. The instructions suggest that subjects be permitted to take the actual depression questionnaire only if they correctly complete 10 of the 15 pretest (Part II) items. However, administrators of the SRDQ are permitted to question the subject further to see if a given item scored as "incorrect" for him or her actually may have been correct. For example, "You sleep in a bed" need not necessarily be answered *most of the time* for some people.

The actual questionnaire comprises 32 items, 31 of which describe depression symptomatology, and these are scored in the same way as the pretest items. Each item is read aloud twice to the person, and a response of *almost never* is allotted a value of 1, *sometimes* 2, and *most of the time* 3. Respondents are asked to rate their feelings over the last two weeks. Two items are "reverse keyed", which means that the scoring is inverted (i.e., given weights of 3, 2, and 1, respectively). The final item (number 32) asks the individual to select from a group of faces, graded from sad to happy, the one that shows how she or he has been feeling for the past two weeks. The possible scores extend from a low of 32 to a high of 98. Higher scores signify greater severity of depressive symptoms. The authors emphasize, however, that the SRDQ is not intended to render a diagnosis of depression. They argue that the principal use of the instrument is to identify individuals with significant depressive symptomatology so that further evaluation can take place (Reynolds & Baker, 1988).

Although the test booklet is silent on this point, the Questionnaire appears to be appropriate for any person able to understand and respond to the component items. Within the field, this presumably would include most mildly retarded and some moderately retarded individuals. Likewise, the administration booklet does not specify who may administer the SRDQ, but it would appear that any responsible adult, especially if given appropriate training in the use of the questionnaire, could supervise its administration. It is not clear from the administration booklet whether respondents can be tested in small groups or whether individual testing is required.

Critique

Psychometric data for the SRDQ are summarized in Table 2 and Appendix B. The reviewer could locate data from only one study, with a total of 83 adult subjects providing valid test protocols (Reynolds & Baker, 1988). Mean depression scores and standard deviation units were presented for the total group and for males and females separately. Alpha coefficients were derived from two test administrations, and both equalled or exceeded .90, which indicates excellent internal consistency. Item-total correlations ranged from .27 to .68, with a mean of .45, which is moderately high. Over an 11-week interval, test-retest reliability was .63, which is moderately good, especially given the length of the time interval.

Construct validity of the instrument hinges upon its relationship to the DSM-III-R and Research Diagnostic Criteria for Depression. Reynolds and Baker also factor analyzed the SRDQ but failed to interpret or discuss the 10 factors that emerged. No criterion group validity data were presented for the questionnaire. Finally, SRDQ scores were correlated with interview scores obtained with the Hamilton (1960) Depression Scale, and the results indicated moderate levels of congruent validity.

The reviewer could find no psychometric data on the pretest component of the SRDQ.

In summary, the Self-Report Depression Questionnaire is a fairly recently developed instrument, and this is reflected by a relatively small amount of psychometric data. Unfortunately, data are lacking on the pretest, which is not a moot point because pretest performance determines whether or not the questionnaire can be regarded as valid for a given subject. Standardization data are available, although the sample size is quite small. Internal consistency appears to be high, and test-retest reliability appears to be acceptable, although more data would be welcome. Factorial/taxonomic validity hinges largely on the relevance of depressive symptoms in the normal IQ population to individuals with mental retardation. With mildly retarded people, this is not likely to be a problem. There is a modicum of congruent validity data, but more are needed. Thus, this instrument falls into a fairly large group of new and promising assessment tools, but much more data are needed before its appropriate niche can be determined.

References

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23*, 56-62.

Reynolds, W. M. (1989, in preparation). *Self-Report Depression Questionnaire (SRDQ) administration booklet*. Odessa, FL: Psychological Assessment Resources, Inc.

Reynolds, W. M., & Baker, J. A. (1988). Assessment of depression in persons with mental retardation. *American Journal on Mental Retardation, 93*, 93-103.

Strohmer-Prout Behavior Rating Scale
D. C. Strohmer & H. T. Prout, 1989

Point-form Synopsis

Stated Purpose: To identify maladaptive behavior and personality patterns among mildly retarded and borderline IQ adolescents and adults.

Age Range: Fourteen years through adulthood.

Level of Mental Retardation Covered: Borderline intelligence and mild mental retardation.

Raters/Diagnosers: Persons who, in their work capacity, are familiar with the individual to be rated.

Time Required to Complete: Fifteen minutes or less.

Disorders/Dimensions Identified: Twelve subscales as follows: (1) Thought/Behavior Disorder, (2) Verbal Aggression, (3) Physical Aggression, (4) Sexual Maladjustment, (5) Noncompliance, (6) Hyperactivity, (7) Distractibility, (8) Anxiety, (9) Somatic Concern, (10) Withdrawal, (11) Depression, and (12) Low Self-Esteem. In addition, two global factors, Externalizing Factor and Internalizing Factor.

Date of Manual Publication: 1989.

Cost: Complete kit (manual, plus 25 rating sheets and scoring booklets), $72.50; 25 rating sheets, $24.50; 25 scoring booklets, $17.00; complete kit with computer scoring software, $197.50.

Source: Genium Publishing Corporation, Psychological Testing Division, Department PS9A, 1145 Catalyn Street, Schenectady, NY 12303-1836. Telephone (518) 377-8854; FAX (518) 377-1891.

Limitations/Exclusions: Adolescents and adults with moderate through profound mental retardation; children less than 14 years of age; not normed for ratings by parents or guardians.

Description

The Strohmer-Prout Behavior Rating Scale (SPBRS) is a 135-item scale for rating the behavior of adolescents (14 years of age and older) and adults with borderline intelligence or mild mental retardation (IQ 55-83). It was developed using a "rational/clinical" method, in which the component subscales and their respective items were determined by the authors in consultation with experienced workers in the field. This was followed by correlational and confirmatory factor analysis in which an attempt was made to validate the structure of the scale.

The 12 subscales of the SPBRS have been designated as follows: (1) Thought/Behavior Disorder (15 items), (2) Verbal Aggression (8 items), (3) Physical Aggression (10 items), (4) Sexual Maladjustment (8 items), (5) Noncompliance (15 items), (6) Hyperactivity (10 items), (7) Distractibility (10 items), (8) Anxiety (11 items), (9) Somatic Concerns, (12 items), (10) Withdrawal (10 items), (11) Depression (11 items), and (12) Low Self-Esteem (15 items). In addition, separate Externalizing and Internalizing factors are calculated. The Externalizing Factor is determined by adding the raw scores from the Verbal Aggression, Physical Aggression, Noncompliance, and Hyperactivity subscales. The Internalizing Factor is computed from the sum of the Anxiety, Depression, and Low Self-Esteem subscales.

Informants are intended to be caregivers who are familiar with the individual being rated, such as rehabilitation counselors, work supervisors, teachers, vocational evaluators, residence counselors, psychologists, and so forth. The instrument is not intended to be completed by parents or guardians, as relevant norms do not exist for parent-figures. The manual reports that completion time typically is 15 minutes or less per individual, although it took the reviewer slightly longer to rate a hypothetical person. Normative data are based on samples taken from a variety of day programs (ranging from institutional through competitive employment) and residential programs (ranging from developmental centers through independent living). The developers of this instrument espouse a multi-method approach to the assessment of social-emotional and behavioral problems in persons with mental retardation. To this end, they recommend employing other clinical information such as observational and interview data, as well as the individual's self ratings when using the

SPBRS. A companion instrument for obtaining self ratings, the Prout-Strohmer Personality Inventory, is reviewed elsewhere in this report.

Additional Features

Software is available (for IBM compatible PCs) for scoring the SPBRS and for providing a graphic display and clinical interpretation of the results.

Critique

The data relevant to the psychometric characteristics of the SPBRS are presented in Table 2 and Appendix B. More than the usual care appears to have been exercised in compiling the normative group for this instrument. As noted above, the rated individuals were sampled from a variety of day and residential programs, and the manual presents a breakdown for major demographic variables such as age, gender, race, and so forth. The manual also contains tables for converting raw scores to percentiles, which is a useful feature.

In terms of reliability, the alpha coefficients for this instrument are all very high, indicating excellent internal consistency. Likewise, item-total correlations were very high, with an overall mean of .71 between individual items and their respective subscales. No test-retest reliability data are presented in the manual, which is surprising given the relatively thorough job in assessing the instrument's psychometric characteristics otherwise. With the exception of the Sexual Maladjustment subscale, data on the instrument's interrater reliability were uniformly very high, with overall means of .82 and .78 obtained across subscales in two separate studies.

Following a review of the relevant literature and interviews with workers in the field, the instrument's developers constructed the subscales and their respective items on an *a priori* basis. A further group of 12 workers with expertise in mental retardation rated the suitability of each item with respect to its underlying dimension. This provides some evidence for the instrument's content validity. Determination of the subscales and their items was based on what the developers characterize as a "rational/clinical" approach. However, it is not clear what diagnostic system or clinical model guided that approach. The instrument's division into 12 subscales was reported to be validated by confirmatory factor analysis. However, no parameters were reported for this procedure, and factor loadings were not tabulated. Thus, the evidence for factorial validity of this scale is unclear at this time. To assess criterion group validity, data were compared for several index groups, such as subjects taking psychotropic drugs (versus those not taking medication), subjects who had a behavior plan to reduce problem behavior, subjects with a DSM-III diagnosis, and so forth. Almost all comparisons of index and non-index groups showed differences in the predicted direction, with index subjects exhibiting more behavior problems. However, there was a frustrating absence of inferential statistics to show exactly which comparisons differed significantly. Finally, a substantial amount of data was offered to demonstrate congruent validity for the scale. There was good correspondence between SPBRS subscale scores and analogous subscales of the Child Behavior Checklist (Achenbach & Edelbrock, 1979). Likewise, SPBRS subscales were moderately-to-strongly correlated with maladaptive subscales on the AAMD Adaptive Behavior Scale (Nihira, Foster, Shellhaas, & Leland, 1975) and the Inventory for Client and Agency Planning. However, subscale scores were only weakly correlated with similar subscales on a self-report companion instrument, the Prout-Strohmer Personality Inventory (Prout & Strohmer, 1989).

To sum up, the standardization for this instrument appears to have been well done with some attempt having been made to include representative samples of mentally retarded individuals. The available data also suggest that the scale has good reliability, although information on test-retest reliability inexplicably is missing. Unlike most behavior scales in this field, evidence is presented relating to the SPBRS's content validity. However, the rationale for determining the scale's structure is not clear, and the available data on the confirmatory factor analysis do not resolve the matter at this stage. There is good evidence for criterion group and congruent validity. At this stage, the SPBRS appears to be one of the better informant rating scales of problem behavior in persons with mild mental retardation, although its division of problem behavior into 12 subscales may be overly fine-grained.

References

Achenbach, T.M. & Edelbrock, C.S. (1979). The Child Behavior Profile: II. Boys aged 12-16 and girls aged 6-11 and 12-16. *Journal of Consulting and Clinical Psychology*, *47*, 223-233.

Nihira, K., Foster, R., Shellhaas, M., & Leland, H. (1975). *Manual for AAMD Adaptive Behavior Scale*. Washington, DC: American Association on Mental Deficiency.

Prout, H. T., & Strohmer, D. C. (1989). *Prout-Strohmer Personality Inventory manual*. Schenectady, NY: Genium Publishing Corporation.

Strohmer, D.C. & Prout, H.T. (1989). *Strohmer-Prout Behavior Rating Scale manual*. Schenectady, NY: Genium Publishing Corporation.

Vineland Adaptive Behavior Scales
(Maladaptive Behavior Domain)
S. C. Sparrow, D. A. Balla, & D. J. Cicchetti, 1984

Point-form Synopsis

Stated Purpose: To assess adaptive behavior for preparing individual educational, habilitative, or treatment programs. The Maladaptive Behavior Domain calls for the rating of "minor" maladaptive behaviors (Part 1) and "serious" maladaptive behaviors (Part 2).

Age Range: Birth through 18 years inclusive.

Level of Mental Retardation Covered: Mild through profound.

Raters/Diagnosers: Professionals, with advanced training in assessment and test administration, who interview the adult most familiar with the person being rated.

Time Required to Complete (Maladaptive Behavior Domain): Estimated by the reviewer at 5 to 12 minutes.

Disorders/Dimensions Identified: Part 1 comprises 27 "minor" behavior problems and Part 2 encompasses nine more severe behavior problems.

Date of Manual Publication: 1984.

Cost: Vineland Adaptive Behavior Scale: Interview edition starter set, $68.00; survey form manual, $24.00; 25 survey form booklets, $22.00; survey form ASSIST, $104.00; complete Vineland starter set, $85.00.

Source: American Guidance Service, Publisher's Building, Circle Pines, MN 55014-1796. Telephone (800) 328-2560 (except Minnesota; [800] 247-5053).

Limitations/Exclusions: Maladaptive Behavior domain not normed for children under 5 years of age.

Description

The Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984) are a revision of the Vineland Social Maturity Scale developed by Doll (1935, 1965). Three versions make up the Vineland Scales; namely, the survey form, the expanded form, and the classroom edition. The survey form contains 297 items, the expanded form has 577 items, and the classroom edition has 244 items. The survey form and expanded form were developed for subjects aged 0 to 18 years 11 months, whereas the classroom edition was developed for children aged 3 years through 12 years 11 months. All three versions of the Adaptive Behavior Scales render four domain scores, intended to reflect aspects of adaptive behavior, as follows: (1) Communication domain, (2) Daily Living Skills domain, (3) Socialization domain, and (4) Motor Skills domain. The authors define adaptive behavior as the performance of the daily activities required for personal and social sufficiency (Sparrow et al., 1984). Both the survey form and the expanded form of the Adaptive Behavior Scale include a Maladaptive Behavior domain, the administration of which is optional.

The Maladaptive Behavior domain was developed only for individuals 5 years of age and older. This domain is composed of two parts. Part 1 contains what the authors characterize as "minor" maladaptive behaviors, and norms are available from both a large national standardization sample and supplementary norm groups. Part 2 describes more serious behaviors, and the pertinent norms are based on the supplementary groups only. Each item on the Maladaptive Behavior domain is rated in terms of frequency as follows: (0) *person never or seldom engages in the activity*, (1) *the person sometimes engages in the activity*, and (2) *the person usually or habitually engages in the behavior*. In addition to frequency, the items in Part 2 are rated for intensity (either moderate or severe). Part 1 of the Maladaptive Behavior domain contains 27 items that describe a heterogeneous collection of behavior problems (e.g., wets bed; bites fingernails; exhibits extreme anxiety). Part 2 contains nine items that are also heterogeneous in terms of constructs assessed (e.g., expresses thoughts that are not sensible; displays behaviors that are self-injurious).

Like the four adaptive behavior domains, Part 1 of the Maladaptive Behavior domain was normed on a national standardization sample totaling 3,000 subjects. Part 2 was normed only for the supplementary groups. Higher scores on the adaptive behavior domain reflect more advanced development. In contrast, higher scores on the Maladaptive

Behavior domain indicate more inappropriate or maladaptive behavior. People who administer the Vineland Scales should be professionals with advanced training in assessment and test administration. Informants can be any adult who knows the individual well, such as parents, house parents, unit aides, social workers, day care workers, and so forth. The time of administration can vary substantially and partially depends on whether Part 2 is given in addition to Part 1. The number of behavior problems exhibited by a given individual also will affect administration time, as the interviewer must probe for frequency and (sometimes) severity data when problems are reported. The principal uses of the Vineland Adaptive Behavior Scales are threefold: (1) to provide diagnostic data, (2) to develop individual educational, habilitation, and treatment programs, and (3) to facilitate research.

Additional Features

Supplementary materials, including a cassette training tape and an Automated System for Scoring and Interpreting Standardized Tests (ASSIST), are available to users. Both English and Spanish versions of the reports to parents are available.

Critique

Of all the instruments reviewed in this report, the Vineland Scales appear to be the most thoroughly standardized, with the national standardization sample carefully stratified on a variety of potentially important background variables. In addition, the availability of the supplementary groups (including ambulatory and nonambulatory mentally retarded institutional residents, mentally retarded adults associated with nonresidential agencies, emotionally disturbed children, visually handicapped children, and hearing-impaired children) is an important feature for professionals working with persons having mental retardation.

Only data relating to the Maladaptive Behavior domain are reviewed here. These are summarized in Table 2 and Appendix B. As determined by split-half reliability coefficients, the internal consistency of the domain appears to be quite high. No item-total correlations were presented for the Maladaptive Behavior domain. Test-retest reliability averaged .88, which is very high, and interrater reliability, although somewhat lower (mean = .74), is still in the high range.

No data could be located relating to the factorial/taxonomic validity of the domain. There are criterion group validity data indicating that emotionally disturbed children scored higher (worse) than ambulatory mentally retarded adults who, in turn, scored higher than the national standardization sample. Likewise, autistic children scored higher than nonautistic, developmentally disabled children (Volkmar, Sparrow, Goudreau, Cicchetti, Paul, & Cohen, 1987). No congruent validity data could be located for the domain.

In summary, the Maladaptive Behavior domain appears to have satisfactory to good reliability levels. However, the lack of data on its validity, especially with respect to factorial validity, is cause for concern. As noted previously, the items comprising this domain are very heterogeneous and appear to address a multitude of types of maladaptive behavior. As such, the principal use for this subscale would appear to be for screening rather than for determination of specific types of aberrant behavior. The subscale warrants inclusion in this review because of the vast popularity of the Vineland Adaptive Behavior Scale, which is one of the most widely used adaptive behavior scales in the mental retardation field. However, the applications of the Maladaptive Behavior domain would appear to be rather narrow, especially insofar as diagnosis of different types of mental disorders are concerned.

References

Doll, E. A. (1935). A genetic scale of social maturity. *American Journal of Orthopsychiatry, 5*, 180-188.

Doll, E. A. (1965). *Vineland Social Maturity Scale.* Circle Pines, MN: American Guidance Service.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Interview edition. Expanded form manual. Vineland Adaptive Behavior Scales.* Circle Pines, MN: American Guidance Service.

Volkmar, F. R., Sparrow, S. S., Goudreau, D., Cicchetti, D. V., Paul, R., & Cohen, D. J. (1987). Social deficits in autism: An operational approach using the Vineland Adaptive Behavior Scales. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 156-161.

Part II

Brief Summaries:
Unpublished and/or Less Established Instruments

Attentional Checklist
J. P. Das & L. Melnyk, 1989

The Attention Checklist is a 12-item scale designed to detect attentional deficits without reference to hyperactive (i.e., overactive) behavior. Although the published report is silent on this, construction of the Attention Checklist appears to be based on the symptoms of the Attention-Deficit Hyperactivity Disorder in the DSM-III-R (Das & Melnyk, 1989). Each item of the checklist is scored on a scale ranging from (1) *Not at all* to (4) *Very much*. Half of the questions are phrased positively and half negatively so that rater responses have to be recoded to reflect the direction of the items. Better attention is signified by higher scores for all items, and possible total scores range from 12 to 48.

The published report of the checklist was based on a study of 100 mildly retarded adolescents who attended a specialized junior/senior high school (Grades 7 to 10) and who were rated by their teachers (Das & Melnyk, 1989). Internal consistency, as assessed by coefficient alpha, was .96 (considered very high), and checklist scores were highly correlated with scores on Conners' (1973) Abbreviated Teacher Rating Scale. A factor analysis of the checklist, using a principal component analysis, rendered one factor, as expected, which explained 71% of the variance.

Thus far, there is very little psychometric information on the checklist, although existing data are very positive. The scale items appear to be internally consistent, to load on one factor, and to correlate well with an established index of hyperactivity. Given the high prevalence and importance of Attention-Deficit Hyperactivity Disorder in children with mental retardation, there could be considerable interest in this tool. At the same time, however, its value with functional levels other than mild retardation is unknown, and very few psychometric data are available thus far.

Source: J.P. Das, Ph.D., Developmental Disabilities Center, 6-123c Education North, University of Alberta, Edmonton, Alberta, Canada, T6G 2G5. Telephone (403) 432-4439.

References

Das, J.P., & Melnyk, L. (1989). Attention Checklist: A rating scale for mildly mentally handicapped adolescents. *Psychological Reports*, *64*, 1267-1274.

Conners, C.K. (1973). Rating scales for use in drug studies with children. *Psychopharmacology Bulletin* (Special Issue on Pharmacotherapy), *126*, 24-28.

Behavior Development Survey
Neuropsychiatric Institute Research Group
at Lanterman State Hospital, 1979

In terms of its data base, this instrument warrants coverage in Part I but it is included here because, according to one of its developers (R. Eyman, personal communication), it has been superseded by the AAMD Adaptive Behavior Scale and the Client Development Evaluation Report. The Behavior Development Survey (BDS) is a behavior assessment instrument designed to assess the adaptive behaviors of developmentally disabled people. The BDS is a modification and briefer version of the AAMD Adaptive Behavior Scale, which is reviewed elsewhere in this report. The BDS renders two types of adaptive behavior summaries. The first part relates to day-to-day adaptive skills and is based on a factor analysis conducted by Nihira (1969) on the domains comprising Part I of the Adaptive Behavior Scale. The three factor scores of the adaptive domains are designated as (1) Personal Self-Sufficiency, (2) Community Self-Sufficiency, and (3) Personal Social Responsibility, and higher scores on these reflect higher levels of adaptive behavior. Two factor scores are derived from the maladaptive behavior section of the BDS. These have been designated as (4) Social Adaptation and (5) Personal Adaptation. The Maladaptive Behavior section of the BDS comprises 11 items related to behavioral and emotional problems. Unlike the Adaptive Behavior Scale, higher scores reflect good rather than poor adaptation. The BDS also contains 19 items not scored onto any of the five factor scores. Each is considered significant in and of itself. These have been divided into four major categories as follows: (1) Health and Medical, (2) Cognitive and Communicative, (3) Social Living, and (4) Personal Problems Requiring Special Attention.

The BDS may be completed by trained professionals or by adults without special training who know the subject well. The survey can be used with both institutionalized and noninstitutionalized subjects with mild through profound mental retardation. Norms are available for ages 6 years through adulthood for institutionalized subjects, whereas they are provided for ages 0 through adulthood for noninstitutionalized subjects. The maladaptive behavior items of the BDS require about 3 to 5 minutes to complete. The major uses for the BDS as stated in the user's manual are twofold: (1) for individual client planning and (2) for administrative planning and evaluation. The BDS can be scored either by hand or by computer. Hand-scored forms result in summary scores which then are converted to percentile scores. The computer-scored alternative produces histogram summaries which are presented as (1) raw scores, (2) percentage of the total score possible, and (3) percentile scores.

The user's manual for the BDS does present interrater reliability levels, which are moderate in size, for the Maladaptive Behavior domains. It also contains extensive normative data on 13,000 institutionalized and 6,000 noninstitutionalized subjects, which are partitioned both by age and by level of mental retardation. Pawlarczyk and Schumacher (1983) assessed the concurrent validity of the BDS with the Vineland Social Maturity Scale, the Peabody Picture Vocabulary Test (PPVT), and the AAMD Adaptive Behavior Scale (ABS). In general, the maladaptive portion of the BDS correlated in the predicted direction with Part II (maladaptive behavior) domains of the ABS. However, the Personal Adaptation domain correlated equally highly with several Part I (Adaptive) domains of the ABS and Socialization on the Vineland Social Maturity Scale, suggesting questionable discriminant validity for this domain. Correlations of the BDS Maladaptive domains with PPVT mental age were low, suggesting independence of ratings from IQ.

Thus, further psychometric studies of the BDS appear to be warranted. The two factors making up the maladaptive portion of the BDS comprise very heterogeneous behavioral items, even within a given factor. Therefore, these dimensions may be useful for screening purposes, but it is unlikely that they would have much utility for establishing the presence of specific emotional or behavioral disorders for research purposes.

Source: Richard Eyman, Ph.D., UC Riverside Neuropsychiatric Institute Research Group at Lanterman Developmental Center, 3530 W. Pomona Boulevard, P.O. Box 100-R, Pomona, CA 91769. Telephone (714) 595-2011.

References

Neuropsychiatric Institute Research Group (1979). *Behavior Development Survey user's manual.* University of California at Los Angeles: Author.

Nihira, K. (1969). Factorial dimensions of adaptive behavior in mentally retarded children and adolescents. *American Journal of Mental Deficiency, 73*, 868-878.

Pawlarczyk, D., & Schumacher, K. (1983). Concurrent validity of the Behavior Development Survey. *American Journal of Mental Deficiency, 98*, 619-626.

Behavior Evaluation Rating Scale
R. L. Sprague, 1982

The *Be*havior *E*valuation *R*ating *S*cale (BeERS) is a 15-item scale designed to measure the effects of medication on problem behavior (R. L. Sprague, personal communication, June 6, 1990). The BeERS is intended for the assessment of adolescents and adults having mild to severe mental retardation. Each item is rated by direct caregivers on a scale that ranges from (0) *Not at all* to (3) *Always*. Except for one item (namely, "Complies with directions–requests"), higher scores on all items reflect worse behavior. Examples of the remaining items include the following: number 3, "Inappropriate verbal behavior;" number 9, "Destructive behavior;" and number 13, "Stereotypic body movements." The scale is largely directed toward the assessment of acting-out, self-injurious, and stereotypic behavior. All scale items are totaled to reflect the global picture for the individual being rated. Ratings are recorded directly onto computer optical scan sheets which also enquire about additional information, such as patient identification, date of rating, and rater identification.

Information is available on the BeERS for a group of 88 residents of a developmental center who were assessed repeatedly by 10 raters (R. L. Sprague, personal communication, June 6, 1990). Frequency distribution data are available for each item broken down for each point on the 4-point scale. In addition, measures of central tendency (mean, standard deviation, skewness, and kurtosis) are available for the 15 items. This, of course, provides a form of standardization for the scale. The reviewer was unable to locate data on reliability or validity with the BeERS. At the present stage of development, this instrument may be helpful for assessing the effects of psychotropic medication, particularly where effects on aggressive and destructive behavior are central issues. The BeERS is probably too narrow in scope to serve as a general diagnostic tool, but may have utility for screening purposes.

Source: Robert L. Sprague, Ph.D., Institute for Research on Human Development,
University of Illinois, 51 Gerty Drive, Champaign, IL 61820. Telephone
(217) 333-4123.

Reference

Sprague, R. L. (1982). *BeERS (Behavior Evaluation Rating Scale).* Unpublished scale, University of Illinois, Champaign.

Behavior Inventory for Rating Development (BIRD)

S. S. Sparrow & D. V. Cicchetti, 1984

The Behavior Inventory for Rating Development (BIRD) (Sparrow & Cicchetti, 1984) is a tool designed to assess types and levels of adaptive behavior in mentally retarded children, adolescents, and young adults. The BIRD has recently been superseded by the Vineland Adaptive Behavior Scales, which are reviewed in Part I of this report (Sparrow, personal communication, October 1989).

The first version of the BIRD bore a different name, the Behavior Rating Inventory for the Retarded (BRIR). The BRIR was constructed to assess five areas of functioning as follows: (1) Communication, (2) Self-help, (3) Psychomotor Skills, (4) Self-control, and (5) Social Behavior (Sparrow & Cicchetti, 1978; Sparrow & Rescorla, 1978). A factor analysis of the BRIR with 45 institutionalizd, mentally retarded children confirmed the existence of four of these five categories as follows: (1) A cognition factor included items previously on the Communication subscale as well as some items from the Social Behavior and Self-help subscales. (2) A psychomotor factor included items from the Psychomotor Skills and the Self-Help subscales. Finally a social and a control factor each corresponded with the Social Control and Self-help subscales, respectively (Sparrow & Cicchetti, 1978). Sparrow and Cicchetti also assessed reliability between raters on different shifts and found that agreement for 6 of 7 and 8 of 10 items exceeded chance levels for the Self-control and the Social Behavior subscales, respectively. Based on work with the BRIR, Sparrow and Cicchetti (1984) subsequently developed the BIRD, which has 75 items. The items are grouped into seven domains, five of which are similar to those in the BRIR: (1) Communication (19 items), (2) Physical Skills (15 items), (3) Self-help Skills (18 items), (4) Self-control (9 items), (5) Social Behavior (8 items), (6) Prevocational Skills (2 items), and (7) Recreational Skills (4 items). Each item is scaled ordinally from a low level of adaptive behavior through to normal behavior, and the ordinal scales use from four to six steps. Teachers in public educational facilities for mentally retarded persons rated 464 children and young adults on the BIRD. Interrater reliability data were available for 403 students, and the median reliability levels using intraclass correlation coefficients, for items in the Self-control and Social Behavior domains were .59 and .58, respectively. Coefficients (r_I s) for the full domains were .81 and .72, respectively. The data for the full sample of 464 subjects were factor analyzed with similar results to the earlier study (Sparrow & Cicchetti, 1978). With respect to the two behavioral domains, five of seven items (71%) from the Social Skills domain emerged on a Social factor, and four of nine Self-control items (44%) landed on an analogous factor.

As noted by the developers of this instrument, norms are not available for the BIRD, at least when it was first reported (Sparrow & Cicchetti, 1984). Inevitably this will detract from its appeal, at least in some research contexts. Thus, findings with the BIRD and its predecessor have been consistent in rendering a four-factor solution that is largely, although not entirely, consistent with the *a priori* placement of its items. It is not clear why the domains were not subsequently realigned to be consistent with the results of the factor analyses. Furthermore, the domains that are related to maladaptive behavior tend to be rather generic in that they subsume a variety of aberrant behaviors under one heading. As such, they may be quite useful for screening purposes, but their utility for diagnosis-specific research is likely to be limited. Available data on the inventory's reliability suggest that the interrater reliability levels for the various domains is quite good.

Source: Sarah S. Sparrow, Ph.D., Child Study Center, PO Box 3333, Yale University,
333 Cedar Street, New Haven, CT 06510. Telephone (203) 785-6227.

References

Sparrow, S.S., & Cicchetti, D.V. (1978). Behavior rating inventory for moderately, severely, and profoundly retarded persons. *American Journal of Mental Deficiency, 82*, 365-374.

Sparrow, S.S., & Cicchetti, D.V. (1984). The Behavior Inventory for Rating Development (BIRD): Assessments of reliability and factorial validity. *Applied Research in Mental Retardation, 5*, 219-231.

Sparrow, S.S., & Rescorla, L.A. (1978). *The Behavior Rating Inventory for the Retarded (BRIR): Manual of instructions for administration and scoring.* Unpublished manuscript, Yale Child Study Center, Yale University, New Haven, CT.

Behavior Problems Inventory
J. Rojahn, 1989

The Behavior Problems Inventory (BPI) is an informant instrument that was designed primarily to assess the prevalence and seriousness of self-injurious and stereotypic behavior. The BPI initially was adapted from a ward observation system developed by Schroeder, Schroeder, Smith, and Dalldorf (1978) and has since undergone several modifications (e.g., Mulick, Dura, Rasnake, & Wisniewski, 1988; Rojahn, 1984, 1986; Rojahn, Fenzau, & Hauschild, 1985). It has been used as part of a nationwide survey in West Germany (Rojahn, 1986) and a community survey conducted in Texas (Griffin et al., 1987). In earlier versions of the scale, raters were asked to rate both frequency and intensity of self-injurious behavior as well as frequency and duration of stereotypic behaviors. In each case, however, the two dimensions were so strongly correlated that they were regarded as largely redundant, and the definitive scale requests only frequency ratings.

The current version of the BPI comprises 15 self-injurious behavior items, five stereotypic behavior items, and nine aggressive behavior items for a total of 29. Assignment of items to sections appears to have been on *a priori* clinical grounds. Information also is requested about demographic characteristics of the subject and relationship of the informant to the subject. The instructions ask the rater to determine whether a given behavioral item applies to the rated individual and, if it does, to rate its frequency on a 6-point scale ranging from (1) *behavior occurs less than monthly* through (6) *occurs more than once per hour*. Each behavioral item is accompanied by a brief definition in terms of observable behavior. The scale appears to be suitable across the full range of mental retardation and to be appropriate for both children and adults. The reviewer estimates that it takes approximately 4 to 7 minutes to complete the background/description sections of the BPI and about 5 to 10 minutes to fill in the rating portions.

There is a modest amount of psychometric data on the BPI. Reliability has been reported to range from somewhat poor (Rojahn, 1984) to mixed (Rojahn, Polster, & Mulick, in press) to very high levels (Mulick et al., 1988; Rojahn, 1986). A cluster analysis also has been reported with the BPI, suggesting that self-injurious behavior may fall into three subtypes. The Behavior Problems Inventory is worthy of consideration for investigators interested in the assessment of self-injurious and stereotypic behavior, although it is probably too narrow in focus to be of general use in the assessment of psychopathology.

Source: Johannes Rojahn, Ph.D., The Nisonger Center for Mental Retardation and Developmental Disabilities, The Ohio State University, 1581 Dodd Drive, Columbus, OH 43210-1296. Telephone (614) 292-9670.

References

Griffin, J. C., Richetts, R. W., Williams, D. E., Locke, B. J., Altmeyer, B. K. , & Stark, M. T. (1987). A community survey of self-injurious behavior among developmentally disabled children and adolescents. *Hospital and Community Psychiatry, 38*, 959-963.

Mulick, J. A., Dura, J. R., Rasnake, J. K., & Wisniewski, J. (1988). *Self-injurious behavior and stereotypy in nonambulatory children with profound mental retardation.* Poster presented at the 2nd Annual North Coast Regional Conference, Society of Pediatric Psychology, Worthington, OH, May 12-14.

Rojahn, J. (1984). Self-injurious behavior in institutionalized, severely/profoundly mentally retarded people: Prevalence and staff agreement. *Journal of Behavioral Assessment, 6*, 13-27.

Rojahn, J. (1986). Self-injurious and stereotypic behavior of noninstitutionalized mentally retarded people: Prevalence and classification. *American Journal of Mental Deficiency, 91*, 268-276.

Rojahn, J. (1989). *Behavior Problems Inventory interview form.* Unpublished instrument, The Ohio State University, Columbus.

Rojahn , J., Fenzau, B., & Hauschild, D. (1985). Selbstverletzungsverhalten bei geistig Behinderten [Self-injurious behavior in the mentally retarded]. *Geistige Behinderung, 24*, 183-192.

Rojahn, J., Polster, L. M., & Mulick, J. A. (in press). Reliability of the Behavior Problems Inventory. *Journal of the Multihandicapped Person.*

Schroeder, S. R., Schroeder, C., Smith, B., & Dalldorf, J. (1978). Prevalence of self-injurious behavior in a large state facility for the retarded. *Journal of Autism and Childhood Schizophrenia, 8*, 261-269.

Communication Style Questionnaire
I. Leudar, 1984

The Communication Style Questionnaire is an informant scale used to reflect the extent to which retarded persons use the maxims of communication in their everyday interactions. Maxims are basically a set of rules that govern the use of language, and they can constrain or expand the meaning of utterances. According to Leudar and Fraser (1985), the understanding of such maxims is important, because different behavior disturbances appear to be associated with violations of different subsets of communicative maxims. Furthermore, work in the linguistics field stresses that understanding language involves interpreting what is said against the background of what those who are interacting already know and assume they know about each other. This, in turn, is determined in large part by the previous use of communicative maxims between the respective communicators.

The Communication Style Questionnaire is a 110-item instrument completed by paraprofessionals (e.g., nurses, workshop instructors, etc.) and professionals who know the individual well. Ninety-nine of these items resolve onto 12 subscales, derived by factor analysis, that reflect communication maxims (Leudar, 1989). Each item is rated on a 5-point Likert scale ranging from (0) *never* through (2) *occasionally* to (4) *very often*. The subscales have been designated as follows: (1) Quality, (2) Irrelevance, (3) Quantity, (4) Manner – Prolixity, (5) Manner – Incoherence, (6) Manner – Speech Impediment, (7) Indirectness, (8) Disclosure, (9) Communality, (10) Hostility, (11) Uncooperativeness, and (12) Conflict/Conflict Avoidance. Completion of the Questionnaire is estimated by the reviewer to require approximately 12 to 16 minutes.

Factor solutions for the Communication Style Questionnaire were very similar for mentally retarded and normal IQ samples. However, data also have been presented showing differences in the degree to which maxims of conversation were in power for each of these samples (Leudar, 1989). Most importantly for the present review, Leudar also has published data showing moderately strong to very strong relationships between communicative background and behavior disturbance as assessed by the Behavior Disturbance Scale (reviewed elsewhere in this report). The reviewer was able to find only a modicum of other psychometric data on the instrument. The Questionnaire represents a different approach to assessing behavior/emotional disorders, and its association with ratings of behavior disturbance suggest that this may prove to be a profitable line of investigation (e.g., Leudar, Fraser, & Jeeves, 1987).

Source: Dr. Ivan Leudar, Psychology Department, The University of Manchester, Manchester M13 9PL, England.

References

Leudar, I. (1984). *Communication Style Questionnaire 1*. Unpublished instrument, University of Manchester, Manchester, England.

Leudar, I. (1989). Communicative environments for mentally handicapped people. In M. Beveridge, G. Conti-Ramsden, & I. Leudar (Eds.), *Language and communication in mentally handicapped people* (pp. 274-299). New York: Chapman & Hall.

Leudar, I., & Fraser, W. I. (1985). How to keep quiet: Some withdrawal strategies in mentally handicapped adults. *Journal of Mental Deficiency Research*, *29*, 315-330.

Leudar, I., Fraser, W. I., & Jeeves, M. A. (1987). Theoretical problems and practical solutions to behaviour disorders in retarded people. *Health Bulletin*, *45*, 347-355.

Developmentally Delayed Children's Behaviour Checklist (DDCBCL)-- Primary Carer Version

S. Einfield & B. J. Tonge

(1990)

The Developmentally Delayed Children's Behaviour Checklist (DDCBCL) was still under development at the time of preparing this review. It is designed to be completed by lay people who know the child well. The instrument is intended to be suitable for youngsters with "moderate" and "severe" mental retardation and residing either in the community or in residential settings. The age range for which the scale is suited is not specified, although the initial report indicates that it will be used to assess both children and adolescents.

The DDCBCL is made up of 91 behavioral items, plus four additional slots where further behavior problems can be added by the rater. Each item is rated on a 3-point scale ranging from (0) *not true (as far as you know)* to (2) *very true or often true*. Items were developed by examining 700 clinical files for descriptors of behavior problems and rewriting these for use in the scale. The research plan calls for the DDCBCL to be analyzed by factor analysis to derive appropriate subscales. Plans are in place to examine both interrater and test-retest reliability of the instrument (Einfield & Tonge, 1990). Preliminary analyses with small groups of subjects have produced the following interrater reliability results: between parents (r=.74, N=18), residential care workers (r=.68, N=15), and residential nurses (r=.41, N=33) (S. Einfeld, personal communication, May 2, 1990). It was not clear from this communication whether these were total scale, subscale, or item reliabilities. Validity will be addressed by comparing scale scores for subjects having emotional/behavioral disturbances with those for children free of significant emotional or behavioral disturbance. Research plans also call for congruent validity to be assessed by comparing derived scores with ratings on several other adaptive and maladaptive behavior instruments (Einfield & Tonge, 1990). The reviewer estimates that it would take approximately 10 to 15 minutes to complete the preliminary version of the DDCBCL, although the developed instrument may well prove to be briefer. Obviously, no conclusions can be drawn about the instrument's psychometric characteristics at this time.

Source: Dr. Stewart Einfield, Department of Child, Adolescent and Family Psychiatry, The Children's Hospital, Camperdown, New South Wales 2050, Australia. Telephone (02) 692-6561 or (02) 692-6562. FAX (02) 692-4203.

Reference

Einfield, S.L., & Tonge, B.J. (1990). *Development of an instrument to measure psychopathology in mentally retarded children and adolescents.* Unpublished manuscript, University of Sydney, Sydney, Australia.

Fairview Maladaptive Behavior Survey
J. Barron, 1981

The Fairview Maladaptive Behavior Survey is an interview/informant instrument comprising 206 maladaptive or inappropriate behavioral items sometimes observed in mentally retarded individuals. The items are grouped according to the major areas of maladaption as defined by the State of California as follows: (1) Harm to Others, (2) Harm to Self, (3) Harm to Physical Environment, (4) Inappropriate Activity Level, and (5) Socially Undesirable Behavior. The survey is intended for use with all age levels, all degress of mental retardation, and for individuals residing both in institutional and noninstitutional settings. Time of administration can vary widely depending upon the nature and severity of the behavior problems encountered, and it can range from a few minutes to 1 1/2 hours. The stated purposes of the survey are (1) to assess the behavioral readiness of institutional residents to progress to less restrictive placements and (2) to help in developing guidelines for modifying a given subject's maladaptive behavior. It is clear, however, that the survey could serve a more descriptive function as well, such as in selecting subjects with particular behavioral characteristics.

The instructions call for a trained examiner to interview the informant(s), who has recent detailed knowledge of the person concerned. The informant is asked to identify behavioral items that he or she personally has observed the client performing. Each item is scored using a 6-point temporal key (H = *At least hourly*, through R = *Rarely [once a year]*). For each item that the informant has observed to occur, he or she also is asked to judge its severity, and management response. Severity is coded with a 9-point scale ranging from (1) *Occurs but no injury/damage results from the behavior* through (9) *Can lead, over a period of time, to a life threatening situation.* Management response is coded to indicate the usual form of management required by staff members to control the person's behavior. This encompasses 20 types of interventions that are nested according to the severity of the intervention into one of four major categories as follows: (1) Positive Behavior Interactions, (2) Mildly Restrictive Procedures, (3) Moderately Restrictive or Aversive Procedures, and (4) Highly Restrictive or Aversive Procedures. There are also codes to indicate the duration of the behavioral item (*less than 1 minute* to *25 minutes or more*), as well as the antecedents that typically precede the behavior (17 categories are provided). The writer could find no psychometric data on the Fairview, which is still in the developmental stage (J. Barron, personal communication, November, 1989).

Source: Jennifer Barron, Ph.D., Fairview Developmental Community, 2501 Harbor Boulevard, Costa Mesa, CA 92626. Telephone (714) 957-5534.

Reference

Barron, J. (1981). *Manual and test for the Fairview Maladaptive Behavior Survey.* Unpublished document, Fairview Developmental Community, Costa Mesa, CA.

Gilson-Levitas Diagnostic Criteria

Modifications for Mildly and Moderately Retarded Adults

S. F. Gilson & A. Levitas, 1988

The Gilson-Levitas Diagnostic Criteria are a set of guidelines for the identification of psychiatric disorders and neurological disorders having behavioral components in mentally retarded adults. The criteria are derived from the DSM-III-R, and the diagnostic descriptions have been rewritten in common language, avoiding technical psychiatric terminology where possible. Frequently, diagnostic descriptions have been supplemented by providing definitions, many of which are taken from *The Mosby Medical Encyclopedia* (Glanze, Anderson, & Anderson, 1985).

The modified criteria were tailored for use by mental retardation workers who predominantly hold bachelor's degrees, such as case managers, rather than for professionals with specialized training in the diagnosis of psychiatric disorders. Judging from the title of the instrument, it appears to have been developed for adults with mild and moderate mental retardation, although one paper indicates that the criteria have been used with at least some severely retarded subjects (Gilson, Levitas, & Mead, 1989). Unlike the DSM-III, the instructions for the Diagnostic Criteria call for symptoms to be scored positive if the person being assessed has *ever* exhibited the characteristics of a given disorder. Also, distinct from the DSM-III-R, no fixed numbers of symptoms are specified within the Diagnostic Criteria for a psychiatric condition to be scored as present (Gilson et al., 1989).

The Diagnostic Criteria are not intended to render a definitive diagnosis for affected individuals. Instead, its stated purposes are (1) to serve as a survey tool for estimating the number of retarded people suffering from an identifiable psychiatric or neurological disorder, (2) to identify individuals requiring further evaluation, and (3) to provide relevant information to those serving this population. The major categories within the Diagnostic Criteria include the following: (1) Psychiatric Disorders (e.g., Schizophrenia, Mood Disorder, Anxiety Disorder); (2) Neurological/Metabolic Disorders, (3) Medication Side Effects, (4) Autistic Disorder, (5) Personality Disorders, (6) "Other" Disorders, and (7) Mental Health Problem Not Otherwise Specified. The diagnostic criteria have been applied in one large prevalence study of 5,000 mentally retarded subjects in Colorado (Gilson et al., 1989). There is also a small amount of reliability data available with the instrument (Gilson et al., 1989).

Source: Stephen French Gilson, LCSW, Child Development Center, School of Medicine, Georgetown University, 3800 Reservoir Rd, N.W., Washington, D.C. 20002.

References

Gilson, S. F., & Levitas, A. (1988). *Gilson-Levitas diagnostic criteria: Modifications for mildly and moderately retarded adults.* Unpublished document, Georgetown University, Washington, D.C.

Gilson, S. F., Levitas, A., & Mead, C. (1989). *Psychiatric disorders in community-based mentally retarded adults–A statewide survey.* Manuscript submitted for publication, Georgetown University.

Glanze, W. D., Anderson, K. N., & Anderson, L. E. (Eds.), (1985). *The Mosby Medical Encyclopedia.* New York: C.V. Mosby.

Motivation Assessment Scale
V. M. Durand, 1986

The Motivation Assessment Scale (MAS) approaches "diagnosis" from a very different perspective than other instruments reviewed in this report. Instead of focusing on the form or structure of inappropriate behaviors (e.g., acting-out vs. withdrawn), the MAS instead is designed to assess what purpose is served by the maladaptive behavior. This results in a classification of problematic behaviors according to their presumed communicative functions. Four categories of possible maintaining variables are assumed, namely (1) social attention, (2) tangible consequences, (3) escape from aversive situations, and (4) sensory consequences. The MAS comprises 16 items that are completed by significant others, such as teachers. Each item assesses the likelihood that some specific target behavior will occur in a variety of situations (e.g., following a request to perform a difficult task [escape function]; whenever significant others stop attending to the subject [attention function]; and so forth). Each of the questions is rated on a 7-point Likert scale ranging from (0) *Never* through (3) *Half the time* to (6) *Always*. The scale is structured such that the 16 items resolve into four subgroups of four items each that provide Sensory, Escape, Attention, and Tangible scores. Although the scale description does not actually state this (Durand, 1988), it appears that each target behavior must be rerated for the same set of 16 items. Hence, it is very possible that different behaviors may have completely different functions and thus would be scored differently.

A variety of psychometric data are available on the MAS including interrater and test-retest reliability, which are reported to be high (Durand, 1988; Durand & Crimmins, 1988). However, Sturmey (1989) found the instrument's psychometric properties to be much weaker. Of considerable interest are studies in which the MAS has been used to determine the function of problem behaviors and thereby to help in developing apparently effective behavioral strategies for reducing these behaviors (e.g., Durand & Crimmins, 1988; Durand, Crimmins, Caulfield, & Taylor, 1989; Durand & Kishi, 1987). Thus, this appears to be a potentially useful instrument for suggesting the use of specific types of behavioral habilitative programs. However, the MAS does not appear to have a diagnostic application in the traditional sense; namely, to describe the topographic appearance of several inappropriate and maladaptive behaviors viewed in unison.

Source: V. Mark Durand, Ph.D., Department of Psychology, State University of New York at Albany, 1400 Washington Avenue, Albany, NY 12222. Telephone (518) 442-4845.

References

Durand, V. M. (1986). *Motivation Assessment Scale.* Unpublished instrument, State University of New York at Albany.

Durand, V. M. (1988). The Motivation Assessment Scale. In M. Herson & A. S. Bellack (Eds.) *Dictionary of behavioral assessment techniques* (pp. 309-310). New York: Pergamon Press.

Durand, V. M., & Crimmins, D. B. (1988). Identifying the variables maintaining self-injurious behavior. *Journal of Autism and Developmental Disorders, 18,* 99-117.

Durand, V. M., Crimmins, D. B., Caulfield, M., & Taylor, J. (1989). Reinforcer assessment I: Using problem behavior to select reinforcers. *Journal of the Association for Persons with Severe Handicaps, 114,* 113-126.

Durand, V. M., & Kishi, G. (1987). Reducing severe behavior problems among persons with dual sensory impairments: An evaluation of a technical assistance model. *Journal of the Association for Persons with Severe Handicaps, 12,* 2-10.

Sturmey, P. (1989, September). The Motivation Assessment Scale: Report of poor psychometric properties. In J. Rojahn (Chair), *Behavior Modification Research in Mental Retardation.* Symposium conducted at the meeting of the 19th Annual Congress of the European Association of Behavior Therapy, Vienna.

Psychosocial Behaviour Scale
C. A. Espie. J. M. Montgomery, & J. B. Gillies, 1988

The Psychosocial Behaviour Scale (PBS) is an informant instrument for rating problem behavior in mentally retarded adults. The instrument was developed by factor analysis of the ratings of 130 individuals attending adult training centers. Most of the persons in the developmental groups had either mild or moderate mental retardation.

The PBS comprises 36 items that are rated on a 5-point scale ranging from (0) *behavior never occurs* through (4) *behavior occurs frequently in a stronger/more problematic form.* Twenty-nine of the items resolve onto one or more of the five factors. The factors have been designated as follows: (1) Physical Aggression (7 items), (2) Passivity/Dominance (8 items), (3) Attention-Seeking (7 items), (4) Social Adaptation/Dysfunction (7 items), and (5) Physical Handicap (4 items). Some items score on two subscales, and seven items are not scored on any subscale. Spearman rank order correlations between different subscales were quite high, ranging from .28 to .83 (*M*=.53). Coefficient alpha ranged from .65 to .93 across subscales, with a mean of .81. Item-total correlations were quite high for the various subscales. However, the original report (Espie, Montgomery, & Gillies, 1988) contained no data on reliability or validity. The authors did indicate that further work was in progress, but they did not indicate its nature.

In developing the PBS, the authors particularly were interested in problems related to pseudoseizures (Montgomery & Espie, 1986). As such difficulties are said to be indicative of habitual "hysterical" responses in some individuals, the authors attempted to focus upon behaviors characteristic of an "hysterical response tendency" (e.g., liability to illness, stagy reactions, and manipulative, attention-seeking behavior). The publication describing the PBS does not indicate who may serve as raters, but it would appear that any responsible adult who has good familiarity with the individual could perform such ratings. The PBS is brief and largely untested psychometrically. However, with time it may be found to provide a useful profile for certain purposes, although the tendency of the various subscales to correlate moderately highly with one another suggests that this instrument may assess a somewhat narrow set of behavioral problems.

Source: Dr. Colin A. Espie, "Moorview," Ravenspark Hospital, Irvine KA12 8SS,
Scotland, United Kingdom. Telephone 011-44-294-74191 (Extension 3440).

References

Espie, C. A., Montgomery, J. M., & Gillies, J. B. (1988). The development of a psychosocial behaviour scale for the assessment of mentally handicapped people. *Journal of Mental Deficiency Research*, *32*, 395-403.

Montgomery, J. M., & Espie, C. A. (1986). Behavioural management of hysterical pseudoseizures. *Behavioural Psychotherapy*, *14*, 334-340.

Revised Children's Manifest Anxiety Scale
"What I Think and Feel"
C. R. Reynolds & B. O. Richmond, 1985

The original Children's Manifest Anxiety Scale (CMAS) (Castenada, McCandless, & Palermo, 1956) was a downward extension of a popular manifest anxiety scale developed for use with adults (Taylor, 1953). Since then, the children's version has been revised and published as the Revised Children's Manifest Anxiety Scale (RCMAS) (Reynolds & Richmond, 1985). The scale comprises 37 declarative sentences to which the child must respond Yes or No. Eight of these contribute to a lie score (i.e., the tendency of children to "fake good"), and 31 items contribute to the child's anxiety score. The range of possible scores extends from 0 to 31 on the Total Anxiety portions of the instrument. The revised instrument also has three factor-based subscales, designated as Physiological Anxiety, Worry/Oversensitivity, and Social Concerns/Concentration.

A number of studies were published involving the original CMAS in comparisons of mentally retarded and nonretarded children. In general, these indicated significantly higher scores (suggesting higher anxiety) in the groups of mentally retarded children (Carrier, Orton, & Malpass, 1962; Cochran & Cleland, 1963; Malpass, Mark, & Palermo, 1960; Matthews & Levy, 1961), although not always (e.g., Lipman, 1960). The same is also generally true of lie scale scores. The writer could find only a modest amount of psychometric data involving the performance of mentally retarded children on either the CMAS or the RCMAS. Matthews and Levy (1961) found test-retest correlations of .84 and .86 for the Anxiety Scale and Lie Scale, respectively, in a group of mentally retarded men. However, they also found modest correlations between a specially constructed response set scale and anxiety scores, suggesting some tendency to acquiesce in these subjects. In a study by Pryer and Cassel (1962), subjects were divided into a low mental age (MA) group (6 to 7 years, inclusive) and a high MA group (8-10 years). Test-retest reliability coefficients (r_S) over one week were found to be .63 for the low MA group and .83 for the high MA group. Another study reported correlations between the Prout-Strohmer Personality Inventory (Anxiety subscale) and the RCMAS of .88 for the Total Anxiety score, and.76, .83, and .76 for the Physiological Anxiety, Worry/Oversensitivity, and Social Concerns/Concentration subscales, respectively (Prout & Strohmer, 1989).

Flanigan, Peters, and Conry (1969) adopted a different approach by conducting a statistical item analysis for a group of children with mild mental retardation and controls matched for chronological age. They found that the anxiety scale items did not function in the same way for subjects in the mental retardation and control groups. For example, items associated with higher anxiety for one group were infrequently associated with higher anxiety in the other. This observation raises questions as to whether the scale serves an analogous function in both populations. However, the issue was not resolved in this study, because the authors failed to control for MA. Interestingly, Flanigan et al. found the internal consistency to be higher for the subjects with mental retardation (alpha=.82) than for controls (alpha=.67).

In summary, there is a certain amount of test-retest reliability data with the RCMAS and its predecessor, the CMAS and, in general, these range from adequate to quite good. One study indicated a mild tendency for subjects with mental retardation to acquiesce on the CMAS, although a similar comparison was not carried out to determine whether this also occurred with control subjects (Matthews & Levy, 1961). The item analysis of Flanigan et al. (1969) suggests that the instrument may tend to assess a different construct in the two populations although more systematic study is needed before this conclusion can be accepted with confidence. On the basis of the limited data currently available, the RCMAS and its predecessor appear to be reasonably reliable instruments, but their validity requires much more systematic study in this clinical population.

Source: Western Psychological Services, 12031 Wilshire Blvd., Los Angles, CA 90025.
Telephone (213) 478-2061.

References

Carrier, N. A., Orton, K. D., & Malpass, L. F. (1962). Responses of bright, normal, and EMH children to an orally-administered children's manifest anxiety scale. *Journal of Educational Psychology, 53*, 271-274.

Castenada, A., Palermo, D S., & McCandless, B. R. (1956). The children's form of the Manifest Anxiety Scale. *Child Development, 27*, 317-326.

Cochran, I. L., & Cleland, C. C. (1963). Manifest anxiety of retardates and normals matched as to academic achievement. *American Journal of Mental Deficiency, 67*, 539-542.

Flanigan, P. J., Peters, C. J., & Conry, J. L (1969). Item analysis of the Children's Manifest Anxiety Scale with the retarded. *The Journal of Educational Research, 62*, 472-477.

Lipman, R. S. (1960). Children's manifest anxiety in retardates and approximately equal M.A. normals. *American Journal of Mental Deficiency, 64*, 1027-1028.

Malpass, L. F., Mark, S., & Palermo, D. S. (1960). Responses of retarded children to the CMAS. *Journal of Educational Psychology, 51*, 305-308.

Matthews, C. G., & Levy, L. H. (1961). Response sets and manifest anxiety scores in a retarded population. *Child Development, 32*, 577-584.

Prout, H. T., & Strohmer, D. C. (1989). *Prout-Strohmer Personality Inventory manual.* Schenectady, NY: Genium Publishing Corporation.

Pryer, M. W., & Cassel, R. H. (1962). The Children's Manifest Anxiety Scale: Reliability with aments. *American Journal of Mental Deficiency, 66*, 860.

Reynolds, C. R., & Richmond, B. O. (1985). *Revised Children's Manifest Anxiety Scale manual.* Los Angeles: Western Psychological Services.

Taylor, J. A. (1953). A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology, 48*, 285-290.

Social and Emotional Behavior Inventory

W. Vogel, 1976

The Social and Emotional Behavior Inventory (SEBI) is a rating instrument designed for assessing behavior by using written institutional records rather than live behavior. Raters can be nonprofessional personnel (e.g., research assistants) trained to extract the relevant data. The SEBI comprises 15 items, and its purpose is to describe the social and emotional behaviors of institutional residents and, more specifically, their self-control over emotional expression and ability to relate socially to others. Each item is scored on a 5-point scale that describes various degrees of severity of the trait in fairly descriptive behavior terms. The possible range of scores extends from a low of 15 to a maximum of 75. Reliability of the instrument, based on five independent raters' scoring of 15 records, was reported as .90 (Vogel, Kun, & Meshorer, 1968). However, it should be noted that reliability statistics in this case can be illusory. If institutional records fail to mention existing behavior problems or if the severity of those problems is not accurately specified in written records, it is highly unlikely that the interpretation of those records will be accurate. Thus, as the reliability exercises with this instrument did not attempt to measure rated behavior against actual behavior, it would seem that its true reliability is largely unknown. The SEBI has been found to distinguish between children attending special educational classes and those not attending (Vogel et al., 1968) and to differentiate between the behavior of individuals released from institutional care and those who were retained (Vogel, Kun, & Meshorer, 1969). Furthermore, changes in SEBI scores over time were reported to be negatively related to EEG alpha frequency (Vogel, Kun, Meshorer, Broverman, & Klaiber, 1969).

The Social and Emotional Behavioral Inventory (SEBI) was developed in the late 1960s (Vogel et al., 1968; Vogel, Kun, & Meshorer, 1969; Vogel, Kun, Meshorer, Broverman, & Klaiber, 1969). Given the increased knowledge about different types of behavior disorders and the impact of deinstitutionalization, this instrument has probably been superseded by more refined scales.

Source: William Vogel, Department of Psychology, Worcester State Hospital, Worcester, MA 01613.

References

Vogel, W. (1976). Social and Emotional Behavior Inventory (SEBI). In O.G. Johnson (Ed.), *Tests and measurements in child development: Handbook II* (Vol. 2, pp. 1213-1214). San Francisco: Jossey Bass.

Vogel, W., Kun, K.J., & Meshorer, E. (1968). Changes in adaptive behavior in institutionalized retardates in response to environmental enrichment or deprivation. *Journal of Consulting and Clinical Psychology, 32*, 76-82.

Vogel, W., Kun, K.J., & Meshorer, E. (1969). Determinants of institutional release and prognosis in retardates. *Journal of Abnormal Psychology, 74*, 685-692.

Vogel, W., Kun, K.J., Meshorer, E., Broverman, D.M., & Klaiber, E.L. (1969). The behavioral significance of EEG abnormality in mental defectives. *American Journal of Mental Deficiency, 74*, 62-68.

Social Judgment Scale
P. A. Spragg, 1983

The Social Judgment Scale (SJS) is an open-ended test designed to assess an individual's ability verbally to express adaptive social responses when presented with a variety of hypothetical, emotionally reactive situations. The SJS was designed for evaluating subjects with mild and moderate mental retardation, and each item is read aloud to the subject. The scale comprises 22 questions which are intended to suggest or evoke four emotional states as follows: anger, fear (anxiety), gladness, and sadness. Each item is introduced with, "What would you do if..." and then completed with the respective content for that item (such as, "somebody bumps you in the street and doesn't even apologize"?). Each response is scored (0), (1), or (2) depending on the *quality* of the response, its *appropriateness* within the context of the hypothetical situation presented, and its *congruence* (consistency) with the affective content of the item (i.e., reflecting anger, fear, gladness, or sadness). Thus, the range of possible scores extends from a low of 0 through a high of 44. All items are recorded verbatim on the record form provided, and a set of criteria is provided to assist the examiner in scoring each item. The manual recommends that examiners be limited to persons with training in administering norm-referenced tests and experience with mentally retarded persons. Administration time for the SJS is about 15 minutes.

Considerable psychometric data are presented on the SJS, although the pool of subjects tested appears to be rather small. The manual provides standardization data (sample size=48) and internal consistency, reliability, and validity data. The reviewer felt that he might experience some difficulty scoring responses reliably, but reliability data provided on scoring procedures were, in fact, very high. The SJS is similar in format to the comprehension subtests of the Wechsler (1981) Adult Intelligence Scale. As might be expected, because of the test's reliance on the person's ability to express his or her reactions to hypothetical social situations, SJS scores tend to be moderately highly correlated with IQ. The manual emphasizes that the SJS was developed for research, rather than clinical, purposes. However, several potential clinical applications are suggested, including (1) the screening of individuals prior to placement in less restrictive settings, (2) assignment of individuals to various types of social skills training programs, and (3) assisting in the evaluation of the dually-diagnosed person.

Source: Paul A. Spragg, Ed. D., John F. Kennedy Child Development Center, Campus Box C234, University of Colorado Health Sciences Center, 4200 East Ninth Avenue, Denver, CO 80262. Telephone (303) 270-8826.

References

Spragg, P. A. (1983). *Social Judgment Scale (SJS) experimental edition*. Unpublished manual, University of Colorado Health Sciences Center, Denver.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale—Revised*. New York: The Psychological Corporation.

Social Participation Rating Scale
S. R. Kay, 1984

The Social Participation Rating Scale is an informant measure of social functioning in schizophrenic and mentally retarded populations. The scale is made up of only one item, which relates to the individual's physical and emotional involvement in a structured group activity. This is rated on a 6-point scale, ranging from (0) *no participation* through (5) *active/enthusiastic participation*. Each point on the scale is anchored with fairly specific and concrete descriptors specifying the person's level of social involvement. The scale was designed to quantify the manifest social impairment of inpatients in a hospital setting and to help measure progress resulting from treatment. It was specifically designed for assessing schizophrenic adults and mentally retarded residents, especially those thought to have psychoses. The scale was intended for use by mental health professionals and for paraprofessionals without special qualifications following brief training with the tool. According to instructions in the manual, the rated activity always should follow certain standard guidelines, such as (1) taking place in a given activities (meeting) area, (2) occurring without the distraction of foods, drinks, or cigarettes, and (3) structuring of the general discussion or activity by the activities leader. The subjects then are asked to volunteer for various functions within the group, and the interaction is allowed to continue for about 30 minutes. The observer conducts ratings of social participation immediately after the session but away from the subject(s).

Various psychometric data are presented in the manual for this instrument. There are some normative data, although for only 42 mentally retarded subjects. Data also are presented on the scale's reliability (interrater and test-retest), validity (discriminative, congruent, and criterion group validity), and sensitivity to treatment. However, the majority of the available data appears to be based on nonretarded schizophrenic adult samples rather than on mentally retarded samples.

Although this scale may be useful for selecting subjects showing marked social deficiencies, it would not seem to be appropriate as the sole or major assessment instrument for identifying a homogeneous clinical group. However, it may have utility in treatment studies (e.g., Kay, 1980) to evaluate the impact of various therapies in subjects showing defective social relations.

Source: Stanley R. Kay, Ph.D., Bronx Psychiatric Center, New York Office of Mental Health, 1500 Waters Place, Bronx, NY 10461. Telephone (212) 931-0600 (extension 3410 or 3412).

References

Kay, S. R. (1984). Manual for the Social Participation Rating Scale. *Psychological Documents, 14.* (Whole No. 2654).

Kay, S. R. (1980). Disjunctive arousal changes as a consequence of nondrug clinical intervention. *Biological Psychiatry, 16,* 35-46.

Standardized Assessment of Personality

A. H. Mann, R. Jenkins, J. C. Cutting, & P. J. Cowen, 1981

(Adapted by A. H. Reid & B. Ballinger, 1987)

The Standardized Assessment of Personality (SAP) is a semistructured interview, for use with third-party informants, to determine the presence or absence of a personality disorder. The instrument was developed by Mann, Jenkins, Cutting, and Cowen (1981) for psychiatric interviews with a patient's informant to evaluate premorbid personality. The SAP later was adapted by Reid and Ballinger (1987) for use with mildly and moderately retarded institutional residents. In the Reid and Ballinger studies, nurses served as informants, although it is clear than any articulate adult who knows the individual well could serve in this role. The completion of the SAP results in a classification of the subject as normal or into one or more of the following abnormal personality types: (1) Self-conscious, (2) Schizoid, (3) Paranoid, (4) Cyclothymic, (5) Obsessional, (6) Anxious, (7) Neurasthenic, (8) Explosive, (9) Sociopathic, and (10) Hysterical. These personality disorder types conform to the categories in Section 301 of the International Classification of Diseases (ICD-9) except that two categories, Self-conscious and Anxious, have been added to the assessment.

The SAP is divided into three sections. The first section is made up of a general introduction and questions about the relationship of the informant to the patient and their length of acquaintance. In the second section, the interviewer requests a general description of the patient's personality. In the case of psychiatric patients, the informant is asked to focus on an earlier period when the patient was well (Mann et al., 1981). If no indication of personality disorder arises from this, a series of seven standard questions is asked which are relevant to possible abnormal personality types. If these elicit no evidence of abnormality, the interview is terminated, and the individual is classified as normal. However, if the informant uses certain key words (e.g., moody, aggressive, craves attention) in formulating his or her response, then the interviewer determines whether a personality type is present, its prominence (relative to other personality types), and its endurance. For each personality type, two grades are possible: Grade 1 indicates that the personality description matches a category within the SAP but that it is not severe. Grade 2 indicates that the individual is very unusual or handicapped in day-to-day functioning as a result of the personality type.

According to Reid and Ballinger, the Assessment of Personality is suitable for mildly and moderately retarded adults, but it is unlikely to be appropriate for people with severe and profound mental retardation (Ballinger & Reid, 1987; Reid & Ballinger, 1987).

The interview obviously should be used only by professionals having expertise in the area of mental disorders (e.g., psychiatrists and clinical psychologists). The interview is said to require about 10 minutes to complete, but it may tend to take longer when the rated individual has one or more abnormal personality types. Mann et al. (1981) reported moderate interrater reliability levels for four types of personality, with weighted kappa ranging from .60 to .85 (M=.68). "Intertemporal reliability" over one year was reported as very modest in the Mann et al. study — r_S ranged from .13 (Cyclothymic) to .74 (Obsessional) (M=.42). However, the same data, when reinterpreted by weighted kappa, suggested high intertemporal reliability, with kappa ranging from .76 to .96 for three commonly reported personality types (Cutting, Cowen, Mann, & Jenkins, 1986). At this stage there are only limited data with the interview in the mental retardation field. Reid and Ballinger (1987) reported that a large proportion of an institutional sample presented with one or more personality disorders, and Ballinger and Reid (1987) reported moderate to high interdiagnoser reliability using the instrument. The SAP would appear to be a potentially useful tool for assessing personality disorders in mildly and moderately retarded people, although it necessarily would not be sensitive to the diversity of disorders that do not fall under the rubric of personality disorder.

Sources: 1) A. H. Mann, Academic Department of Psychiatry, Royal Free Hospital,
Pond Street, London NW3 2QG, England.
2) Andrew H. Reid, Consultant Psychiatrist, Royal Dundee Liff Hospital,
Dundee DD2 5NF, Scotland. Telephone (0382) 580-441.

References

Ballinger, B. R., & Reid, A. H. (1987). A standardized assessment of personality disorder in mental handicap. *British Journal of Psychiatry, 150*, 108-109.

Cutting, J., Cowen, P. J., Mann, A. H., & Jenkins, R. (1986). Personality and psychosis: Use of the standardized assessment of personality. *Acta Psychiatrica Scandinavica* , *73*, 87-92.

Mann, A. H., Jenkins, R., Cutting, J. C., & Cowen, P. J. (1981). The development and use of a standardized assessment of abnormal personality. *Psychological Medicine, 11*, 839-847.

Reid, A. H., & Ballinger, B. R. (1987). Personality disorder in mental handicap. *Psychological Medicine, 17*, 983-987.

Structured Clinical Interview
P. A. Spragg, 1988

The Structured Clinical Interview (SCI) comprises approximately 130 questions and is intended to complement other types of clinical data by providing information in the cognitive and affective areas. The interview emphasizes a number of DSM-III-R symptoms that are traditionally obtained by self report. The following sections are included within the instrument: (1) Behavioral Observations and Mental Status, (2) Presenting Problem/Chief Complaint, (3) Cognitive-Affective-Behavioral Relationships, (4) Evaluation of Coping Skills, (5) Perception of Self, (6) Interpersonal Functioning, (7) Relationships with Authority, (8) Anxiety Screening, (9) Depression Screening, (10) Psychiatric Screening, (11) Evaluation of Psychosocial Supports, and (12) Summary and Feedback. As a rule, the questions are posed in simple language that is appropriate for persons with mental retardation. The interview generally avoids yes-no questions, relying instead on more open-ended questions and items with choice formats. Several questions are included that are amenable to objective verification, and nine "lie" items also are contained in the interview.

Although the interview booklet does not state this, the SCI would appear to be most suitable for persons of borderline intelligence and mild mental retardation with perhaps some limited application in moderately retarded individuals. The instrument is presumptively suitable only for use by mental health professionals, such as psychiatrists, psychologists, and social workers. According to its developer, the interview is designed largely as a clinical guide, and the interviewer is free to use selected components of the SCI in the context of any alternative interview format.

The SCI appears to be at an early stage of development, and the reviewer is not aware of psychometric data attesting to its utility. The instrument is eight pages in length (double-spaced) and is estimated by the reviewer to take about 30 to 45 minutes to complete.

Source: Paul A. Spragg, Ed. D., John F. Kennedy Child Development Center, Campus Box C234, University of Colorado Health Sciences Center, 4200 East Ninth Avenue, Denver, CO 80262. Telephone (303) 270-8826.

Reference

Spragg, P. A. (1988) *Structured Clinical Interview.* Unpublished instrument, University of Colorado Health Sciences Center, Denver.

Zung Self-Rating Anxiety Scale
W. K. Zung, 1971; Modified for Mildly Mentally Retarded People by W. R. Lindsay & A. M. Michie, 1988

The Zung Self-Rating Anxiety Scale (SAS) was developed by Zung (1971) to assess generalized anxiety and treatment effects in normal IQ clinical populations. The scale consists of 20 questions to which the respondent answers using a 4-point Likert scale which ranges from *None or a little of the time* through *Most of the time.* Each question asks about some aspect of general nervousness or anxiety (e.g., feeling calm) or about physiological manifestations of anxiety (e.g., frequent urination).

Lindsay and Michie (1988) adapted the Zung SAS for use with mildly and moderately mentally retarded adults. To do this, the wording of questions was revised for ease of understanding, and various types of response modes were assessed for reliability. Three types of response alternatives were compared; namely, the standard presentation of response choices (*None of the time* through *Most of the time*), a random presentation of response categories (i.e., items were not always ordered in terms of increasing frequency), and a "yes/no" format. The split-half reliability for the standard response format was only .12 with mentally retarded subjects, whereas the internal consistency (using coefficient alpha) of the random response mode was .58. The yes/no format rendered the highest reliability with a test-retest correlation of .83 over three months and a split-half correlation of .69. Lindsay and Michie (1988) concluded that the only presentation rendering acceptable reliabilities was that which asked the subject to indicate presence or absence of the anxiety symptoms.

The reviewer could locate no validity data with the modified version of the Zung SAS.

Source: Dr. W.R. Lindsay, Tayside Area Clinical Psychology Department, Strathmartine Hospital, Dundee DD3 OPG, Scotland.

References

Lindsay, W.R,. & Michie, A.M. (1988). Adaptation of the Zung Self-Rating Anxiety Scale for people with a mental handicap. *Journal of Mental Deficiency Research, 32*, 485-490.

Zung, W.K. (1971). A rating instrument for anxiety disorders. *Psychosomatics, XII*, 371-379.

Part III

Instruments Not Reviewed But of Related Interest

Cognitive Diagnostic Battery
S. R. Kay, 1982

This battery was designed to differentiate developmental from nondevelopmental (psychiatric) sources of intellectual impairment. The instrument, which includes a psychomotor battery of tests, is intended as an aid in the differential diagnosis of mental retardation and psychosis in adult patients.

Source: Stanley R. Kay, Ph.D., Bronx Psychiatric Center, New York Office of Mental Health, 1500 Waters Place, Bronx, NY 10461. Telephone (212) 931-0600 (extension 3410 or 3412).

Reference

Kay, S. R. (1982). *The Cognitive Diagnostic Battery: Evaluation of intellectual disorders.* Odessa, FL: Psychological Assessment Resources.

Maladaptive Behavior Scale (MABS)
T. I. Thompson, 1988

The Maladaptive Behavior Scale (MABS) is a behavioral scale used to rate one or more target classes of behavior. The scale requires that the rater estimate both the frequency of the behavior and its intensity in a two-dimensional format. Frequency is rated from a low of zero through to more than 12 instances in the previous 8 hours, and intensity is rated from just noticeable through to severe, defined as involving self injury, injury to another person, or property damage. Each rating results in a single point estimate (within a row by column matrix) of each target class of behavior. The MABS provides an overall index of change and can be used to measure treatment effects such as in drug studies.

Source: Travis I. Thompson, Ph.D., Institute for Disabilities Studies, Suite 145, University of Minnesota, 2221 University Avenue Southeast, Minneapolis, MN 55414. Telephone (612) 627-4500.

Reference

Thompson, T. I. (1988). *Maladaptive Behavior Scale (MABS).* Unpublished scale, University of Minnesota, Minneapolis.

Paroxysmal Behavior Scale

L. F. Gourash, W. J. Helsel, & J. Rojahn, 1989

This instrument was developed to assist the clinician in recording the occurrence of behaviors thought to be seizure-related. It can be employed for the initial evaluation of a patient suspected of having seizures and for assessing response to therapy. The Paroxysmal Behavior Scale is made up of two parts, namely a symptom section (10 items) and an intervention section (seven items).

Source: Linda F. Gourash, M. D., Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine, 3811 O'Hara Street, Pittsburgh, PA 15213. Telephone (412) 624-3964.

Reference

Gourash, L. F., Helsel, W. J., & Rojahn, J. (1989). *Paroxysmal Behavior Scale*. Unpublished instrument, University of Pittsburgh, Pittsburgh, PA.

Seizure and Related Behavior Checklist
W. J. Helsel, 1989

Development of this instrument was based on approximately 100 behaviors listed by the International League Against Epilepsy. The scale comprises 42 items and was developed as a screening device for behaviors that may or may not be seizure-related.

Source: William J. Helsel, Ed.D., Psychology Department, Western Carolina Center,
300 Enola Road, Morganton, NC 28655-4608. Telephone (704) 433-2794.

Reference

Helsel, W. J. (1989). *Seizure and related behavior checklist.* Unpublished instrument, University of Pittsburgh, Pittsburgh, PA.

Shortened Stockton Rating Scale
A. H. Pattie & C. J. Gilleard, 1975;
C. J. Gilleard & A. H. Pattie, 1977

The Stockton Geriatric Rating Scale was developed as a measure of behavioral function in geriatric populations (Meers & Baker, 1966). The Shortened Stockton Rating Scale (SSRS) was derived from the original 33-item instrument by reducing it to 18 items on the basis of interrater reliability levels and by eliminating items whose content was not applicable to British populations (Gilleard & Pattie, 1977). The four subscales of the SSRS are designated as (1) Physical Disability (6 items), (2)Apathy (5 items), (3) Communication Difficulties (2 items), and (4) Social Disturbances (5 items), and norms are available for a variety of elderly groups (Gilleard & Pattie, 1977). Results have been presented with elderly nonretarded individuals on the SSRS's concurrent validity with psychiatric diagnosis (i.e., functional vs. organic impairment) (Pattie & Gilleard, 1975) and on its predictive validity over two years (Pattie & Gilleard, 1978). More recently it has been used in elderly mentally retarded patients and found to have moderate interrater reliability and variable reliability (depending upon gender of the patient) with consultant ratings of behavior disorders (Smith, Ballinger, & Presly, 1981).

Source: Anne H.Pattie, M.A., A.B. Ps. S. Principal Clinical Psychologist, Clifton Hospital York, England.
Dr. Anne H.W. Smith, Consultant Psychiatrist, Royal Dundee Liff Hospital, Dundee DD2 5NF, Scotland.

References

Gilleard, C. J. & Pattie, A. H. (1977). The Stockton Geriatric Rating Scale: A shortened version with British normative data. *British Journal of Psychiatry, 131*, 90-94.

Meer, B., & Baker, I. A. (1966). The Stockton Geriatric Rating Scale. *Journal of Gerontology, 21*, 393-403.

Pattie, A. H. & Gilleard, C. J. (1975). A brief psychogeriatric assessment schedule: Validation against psychiatric diagnosis and discharge from hospital. *British Journal of Psychiatry, 127*, 489-493.

Pattie, A. H. & Gilleard, C. J. (1978). The two-year predictive validity of the Clifton Assessment Schedule and the Shortened Stockton Rating Scale. *British Journal of Psychiatry, 133*, 457-460.

Smith, A. H. W., Ballinger, B. R., & Presly, A. S. (1981). The reliability and validity of two assessment scales in the elderly mentally handicapped. *British Journal of Psychiatry, 138*, 15-16.

The Social Performance Survey Schedule
M. R. Lowe & J. R. Cautela, 1978
(Adapted for adults with mental retardation
by J. L. Matson, W. J. Helsel,
A. S. Bellack, & V. Senatore, 1983)

The Social Performance Survey Schedule (SPSS) was developed by Lowe and Cautela (1978) to assess social skills and deficits in adults having normal IQs. The original SPSS was made up of 50 positive and 50 negative items describing social traits, (e.g., self-sacrificing, insensitive), which are defined in fairly specific behavioral terms. Each item is rated on a 5-point scale which ranges from (0) *not at all* to (4) *very much*. The original instrument was designed to be filled in by the person himself or herself. Internal consistency was reported to be high; test-retest reliability was high; and there were modest negative correlations between the SPSS and the Social Avoidance and Distress Scale (Watson & Friend, 1969) suggesting congruent validity.

Matson, Helsel, Bellack, and Senatore (1983) subsequently modified the SPSS for use as an informant scale for rating adults having mild and moderate mental retardation. Items showing poor interrater reliability ($P < .30$) were dropped from the instrument, leaving a 57-item scale. Factor analysis of the modified instrument resulted in four factors as follows: (1) Appropriate Social Skills, (2) Poor Communication Skills, (3) Inappropriate Assertion, and (4) Sociopathic Behavior. It is interesting that the factor analysis separated all positive behaviors into one factor, whereas the negative items were distributed across three factors. Thus, the positive and negative items do not appear to indicate opposite poles of the same dimension. As such, the SPSS actually may assess both social skills and certain forms of maladaptive behavior.

References

Lowe, M. R., & Cautela, J. R. (1978). A self-report measure of social skill. *Behavior Therapy*, *9*, 535-544.

Matson, J. L., Helsel, W. J., Bellack, A. S., & Senatore, V. (1983). Development of a rating scale to assess social skill deficits in mentally retarded adults. *Applied Research in Mental Retardation*, *4*, 399-407.

Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, *33*, 448-457.

Vocational Problem Behavior Inventory
A. M. LaGreca & W. L. Stone, 1982

The Vocational Problem Behavior Inventory (VPBI) is a 48-item checklist designed for assessing problem interpersonal behaviors relevant to sheltered workshop settings. Items are rated on 4-point scales ranging from (0) *Never* through (3) *Regularly*, and higher scores indicate more problems (maximum score, 144). The VPBI is completed by relevant adults, such as teachers and supervisors, on individuals being considered for entry or already placed in workshop settings. For the interested reader, more details are provided in a number of publications (LaGreca & Stone, 1986; LaGreca, Stone, & Bell, 1982, 1983).

Source: Annette M. LaGreca, Ph.D., Department of Psychology, University of Miami, P.O. Box 248185, Coral Gables, FL 33124. Telephone (305) 284-3477.

References

LaGreca, A. M., & Stone, W. L. (1982). *Vocational Problem Behavior Inventory.* Unpublished instrument, University of Miami, Coral Gables, FL.

LaGreca, A. M., & Stone, W. L. (1986). Interpersonal-vocational skill training for the mentally retarded. In R.A. Feldman, & A.R. Stiffman (Eds.), *Advances in adolescent mental health*, (Vol. 2, pp. 87-98). Greenwich, CT: JAI Press.

LaGreca, A. M., Stone, W. L., & Bell, C. R. (1982). Assessing the problematic interpersonal skills of mentally retarded individuals in a vocational setting. *Applied Research in Mental Retardation*, *3*, 37-53.

LaGreca, A. M., Stone, W. L., & Bell, C. R. (1983). Facilitating the vocational-interpersonal skills of mentally retarded individuals. *American Journal of Mental Deficiency*, *88*, 270-278.

Conclusions and Recommendations

Characterization of Existing Scales

In an attempt to bring some order to this literature, the scales reviewed in Parts I and II of this report have been summarized, according to certain key features, in Tables 3 through 7. In Table 3, the available instruments were categorized according to who served as the principal rater or diagnoser. The numbers of instruments completed by the individual being assessed, by informants, or by skilled professionals were 7, 28, and 5, respectively. Thus, informant-type instruments appear to dominate the field at present.

In Table 4, the instruments are classified in terms of the age groups for which they were developed. It can be seen that coverage is best for adults (29 instruments) and worst for children (19 instruments). Instruments that were developed for a special (and usually narrow) clinical purpose (e.g., the Revised Children's Manifest Anxiety Scale) appear in italics in Tables 3 through 7. It is worth noting that a number of instruments tabulated here are adaptive behavior scales and never were intended to be general purpose diagnostic tools. If the special purpose scales and adaptive behavior scales are deleted from Table 4, we find that eight instruments are left for assessing children and 15 for assessing adults. Furthermore, it is the writer's strong impression that the adult scales generally are sounder psychometrically, more comprehensive, and more suitable to their respective populations than are the child scales. We shall return to this point later.

The methods by which the various scales were developed are summarized in Table 5. Ten instruments were empirically derived (usually by factor analysis), 19 were structured using clinical or *a priori* methods, and seven were constructed using a combination of these approaches.

The number of subscales or dimensions for each of the instruments is depicted in Table 6. Twelve instruments have three or fewer subscales. Most of these are either special purpose scales or adaptive behavior scales. One exception is the Preschool Behavior Questionnaire, but it is likely that this scale may fail to assess some commonplace behavioral dimensions. Another exception is the BIRD, which is a predecessor of the Vineland, one of the adaptive behavior scales. The modal number of dimensions falls into the range of five to eight subscales. Instruments having this degree of complexity may have the advantage of providing a reasonable amount of information about the individual without at the same time collecting redundant information (i.e., overlapping subscales) or sacrificing reliability for richness of clinical detail. The instruments having five to eight dimensions appear to come disproportionately from the ranks of empirically-derived or jointly clinically/empirically-derived instruments. However, although it can be argued that the range of five to eight comprises the modal number of subscales (13 of 37 instruments), it is also true that only 35% of available instruments fall into this category. Finally, ten scales have nine or more behavioral dimensions. As noted earlier, it is possible that the structure of some of these (e.g., the Devereux Child Behavior Rating Scale) may be rather unstable due to over-refinement of the factor solution.

Finally, Table 7 contains a summary of this literature in terms of the levels of mental retardation said to be covered by the various instruments. It should be noted that several instruments are claimed to be relevant for assessing a broad range of mental retardation but information attesting to their validity for this is often not available. Thus, this table may suggest a rosier picture than in fact is warranted. Again, specialized instruments appear in italic print. If special purpose scales and adaptive behavior scales are deleted from this summary, we find that there are 16 remaining instruments for assessing mildly retarded people but only ten instruments for evaluating profoundly retarded persons. Again, this appears to indicate another area in need of attention.

State of the Field (Quality of Available Instruments)

Volume of recent work. In assuming the present task, the reviewer had expected to locate relatively few assessment instruments (certainly fewer than 20). In many respects, the number and diversity of available tools is rather surprising. Of the instruments reviewed here, the large majority has appeared since 1984, indicating that instrument development has been a major activity over the last few years. Likewise, there is a range of instruments, such that tools are available for obtaining self ratings, informant reports, and professional assessments. At the same time, however, numbers do not tell the whole story. Many of the instruments reviewed were parts of adaptive behavior scales or were developed for highly specific purposes. Likewise, most of them simply have not been evaluated for their utility as diagnostic instruments.

Comparison with clinical child instruments. It is interesting to compare available reliability data from the scales discussed here with those from the clinical child field reviewed by Achenbach, McConaughy, and Howell (1987). Unfortunately, many of the reliability studies summarized in Appendix B do not allow such a comparison because of incompatible statistics (e.g., the use of *percentage agreement* rather than correlation coefficients). Nevertheless, there are sufficient studies to make a very crude comparison possible. The interrater reliability data from Appendix B were cast by the writer into the same format as employed in Table 1. To achieve this summary, three rules of thumb were adopted. First, where a mean correlation figure was summarized across several subscales of the same instrument, this reliability level (rather than the individual subscale reliabilities) was used to characterize the instrument in question. Second, percentage agreement and unconventional statistics were excluded from this comparison. Third, because of the limited number of studies available, interrater reliabilities from the mental retardation literature were classified without regard to whether the reliability exercise was conducted with similar informants or different types of informants (see Table 1). This classification procedure resulted in 9%, 39%, 17%, and 35% of the reliability coefficients from Appendix B being classified as *poor*, *fair*, *good*, and *excellent*, respectively, using Cicchetti and Sprarrow's (1981) criteria discussed in the Introduction. In all, 52% of the instruments achieved interrater reliability levels characterized as either *good* or *excellent*. Note that the work in the clinical child field resulted in 62% of the obtained interrater reliability correlations falling into the *good* and *excellent* reliability levels for similar types of informants. This figure for the clinical child field drops to only 18% (i.e., 48 of 262 comparisons) if we disregard whether the studies used similar or dissimilar types of informants (see Table 1). This comparison of reliability data across fields is necessarily very crude and subject to all of the caveats alluded to in the Introduction. Nevertheless, by this criterion the available work in the mental retardation field appears quite sound when compared with the psychometric work in the clinical child field, which has considerable sophistication and a substantial background in the use of rating instruments.

Commonalities in factorial studies. One question that naturally arises from this research is whether a common core of behavior disorders emerges from the factor analytic research. Unfortunately, only a few factorially-derived instruments admit to such an exercise. For example, the Devereux scales (Spivack, Haimes, & Spotts, 1967; Spivack & Spotts, 1966) were probably overly refined, and they contain so many subscales that obtaining commonality across instruments is likely to occur simply because of the large number of subscales contained.The development of other instruments (e.g., the Strohmer-Prout Behavior Rating Scale) largely was driven by *a priori* clinical considerations. Another instrument, the Psychosocial Behavior Scale (Espie, Montgomery, & Gillies, 1988) was not included because the range of component behaviors was deemed to be too narrow for comparison with other instruments and the sample size too small for stable factors to emerge. Therefore, the materials adopted for this exercise were as follows. (1) A factor analysis of Part II of the AAMD Adaptive Behavior Scale: Residential and Community Edition by Nihira (1978), which resulted in nine factors. (2) A factor analysis of items within the Aberrant Behavior Checklist by Aman, Singh, Stewart, and Field (1985), which resulted in five factors. (3) Factoring of items comprising the Behaviour Disturbance Scale (Leudar, Fraser, & Jeeves, 1984), leading to a six-factor solution. (4) The Balthazar Scales of Adaptive Behavior-II (BSAB-II), which has seven subscales describing inappropriate behavior (Balthazar & English, 1969). (5) The Diagnostic Assessment for the Severely Handicapped (DASH) Scale, a scale for producing categorical diagnoses, which was also factor analyzed to produce a 6-factor solution (Matson, Coe, Gardner, & Sovner, 1990). (6) The Preschool Behavior Questionnaire, which was developed with children of normal IQ and has a three-factor solution (Behar & Stringfield, 1974).

In Table 8, an attempt has been made to summarize the factor content of these instruments. Five factors appear to emerge with considerable consistency across studies and across instruments. An "Aggressive, Antisocial, and Self-Injurious" factor emerges in part or wholly in all six analyses. It is interesting that self-injurious behavior tended to cluster with unsociable behavior in two of the studies (Aman et al., 1985; Nihira, 1978). In a third study (Matson et al., 1990) some self-injurious behaviors clustered with an Emotional lability factor, also subsumed under the Aggressive behavior heading. However, two other examples of self injury in the same scale failed to cluster with any of the derived factors. There are no self-injurious behaviors on the Preschool Behavior Questionnaire so that such a result is obviated with this instrument. It also is interesting that self-injurious behavior does *not* appear to cluster with other forms of stereotypic behavior, insofar as many workers in the field regard them as closely related clinically. In a nationwide study conducted in West Germany, Rojahn (1986) found that both classes of behavior might co-occur, but there was not a statistical association between them. A variety of other externalizing behaviors also have been arbitrarily placed in the Aggressive, Antisocial, and Self-Injurious category, but it is very possible that future work will disentangle these various forms of acting out onto two or more empirical dimensions.

A "Social Withdrawal" factor (the second category) appears for all five of the more intricate instruments, all of which were derived with mentally retarded populations. Withdrawal-type behaviors also have been noted under category number 6, "Anxious, Tense, and Fearful." Almost all factorial work of this type with children of normal IQ has produced an anxiety/internalizing dimension as the second most prominent factor (Achenbach & Edelbrock, 1978; Quay, 1979). However, in persons with mental retardation, it appears that anxious behavior shows up empirically more as a tendency to withdraw, to be inactive, and to engage in a limited set of activities.

The analyses for four of the scales providing for the inclusion of stereotypic behavior usually resulted in an unequivocal "Stereotypic Behavior" factor. However, this factor was coupled with hyperactive tendencies in one study (Nihira, 1978). In a fifth study, stereotyped movements emerged within the context of a social withdrawal factor (Matson et al., 1990).

Most of the scales provided a dimension that could be encompassed under a separate "Hyperactivity" heading (the fourth category). As noted earlier, hyperactive tendencies were linked with stereotypic behavior in Nihira's investigation. No hyperactivity factor appeared in the analysis conducted by Matson et al. (1990), but there were few elements related to attention deficits and overactivity on their scale and, furthermore, their subjects largely comprised adults who would not be expected to display this behavior pattern nearly as often as children.

"Repetitive Verbalization" shows up as a consistent feature in three of the six instruments. The item content in each of the three cases is fairly similar which, therefore, provides some confidence that this may be a real phenomenon. Several of the factor analyses have rendered dimensions which could be construed as reflecting anxiety and tension, and these comprise the sixth category of the table. Finally, "Self-Injurious Behavior" is entered as a separate category (the seventh category) because at least one study (Leudar et al., 1984) found a separate factor describing self-injurious activity. However, Matson et al. (1990) noted that self injury failed to emerge as a separate category, despite the presence of several relevant items on that scale.

To sum up, there is a moderate, although certainly not unanimous, degree of consistency across these factor analytic studies. The most consistent dimensions appear to be as follows: (1) Aggressive, antisocial, and self-injurious behavior. (However, the several acting-out behaviors subsumed here may require more than one dimension to account for them statistically.) (2) Social withdrawal (perhaps combined with anxiety and mood fluctuations) also appears to be a consistent dimension. (3) Stereotypic behavior shows up as a separate dimension on most instruments making provision for these repetitive activities. (4) Hyperactivity is a common, although not universal, dimension found in factorial studies. Not surprisingly, given what we know from clinical studies with nonretarded populations, this pattern is much more conspicuous in studies using younger subjects. (5) Inappropriate, repetitive vocalizations emerged as a dimension in half of the instruments reviewed. Of course, there could be additional relevant dimensions that did not show up here. In order for behavior problems or symptoms to emerge as part of a separate factor, they must, of course, be present in the item pool. We have no assurance that previous research has included all relevant forms of maladjustment. However, this does provide us with some idea of what the structure of the more commonly occurring maladaptive behaviors may look like.

Recurring Problems with Available Instruments

Despite the fact that this has been an area of considerable research activity, there remain a number of common problems insofar as the relevant literature is concerned. One of these concerns the sensitivity and specificity of the instruments reviewed. *Sensitivity* refers to the probability that a person who has a psychiatric or behavioral disorder will be classified, in fact, by the instrument as having a disorder. *Specificity* is expressed as the probability that a person without a psychiatric or behavioral condition will be classified by the instrument as not having the disorder (Kleinbaum, Kupper, & Morgenstern, 1982). Of all the instruments reviewed here, only one comes to mind (i.e., the Reiss Screen) that presented data showing the correctness of the classification achieved. Of course, part of the problem has been alluded to earlier; namely, the lack of a gold standard for validation purposes. In the case of the Reiss (1988) Screen, the DSM-III was used as the criterion, although the appropriateness of this may be open to debate. Nevertheless, the collection of such data in relation to designating a person as having a disorder or not is a goal well worth striving for in future research. However, the determination of suitable criterion devices may require considerable ingenuity on the part of investigators.

A related but more specific question concerns the diagnostic precision of available instruments. That is, what is the accuracy of the several disorders or dimensions on a given instrument? This is really a more refined or specific variation of the "sensitivity-specificity" question raised immediately above. Thus far, attempts to address this question have been rather piecemeal and apparently contingent on the availability of related instruments from adult psychiatry. For example, some developers of multidimensional instruments have looked at the accuracy of depression or anxiety components in relation to self ratings of depression/anxiety on other instruments, but the accuracy of the remainder of the instrument generally has gone unanswered. Once again, this probably reflects the lack of a gold standard for this field, which is a problem that we, as a scientific community, must surmount.

Another common deficiency in the instruments reviewed relates to inadequate standardization. Only the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984) achieved true standardization, with a large normative sample representative of the U.S. population in terms of sex, race or ethnic group, geographic region, community size, and parents' education. It is probably not necessary to achieve this level of sophistication for a screening or diagnostic tool to be useful and accurate, but it is also clear that salient variables such as age, gender, level of mental retardation, residential setting, and so forth should be taken into account. Most of the instruments reviewed failed to achieve much sophistication in this regard, although a few (such as the Prout-Strohmer Assessment System) (Prout & Strohmer, 1989; Strohmer & Prout, 1989) appeared to be quite adequate.

Thus, to summarize, the major problems with the available instruments as a group is that (1) their sensitivity and specificity largely is unknown, (2) their diagnostic accuracy essentially is untested, and (3) standardization often is inadequate. To a large extent, this is due to a lack of well-accepted validating criteria. Many of these instruments were never intended to be diagnostic tools as such and, hence, their failure to present sensitivity and specificity data is understandable. Finally, most of these instruments were developed or assessed on fairly small budgets, often without any external funding, which helps to explain the lack of extensive standardization data. Nevertheless, these are important standards to look for when we try to locate a suitable instrument for a given research or clinical purpose. Despite the multitude of existing instruments in the field, the numbers can be somewhat deceptive and perhaps falsely reassuring. If we partition the available assessment tools according to their established indications (e.g., in terms of age, level of mental retardation, type of rater, specific psychological/psychiatric conditions covered, and so forth), the range of instruments for any given specific need is greatly reduced. If these tools are required to meet all of the above standards, the number of available instruments becomes very small indeed.

Towards a Valid Taxonomy of Emotional and Behavioral Disorders in Mental Retardation

The measurement and determination of the nature of mental disorders in mental retardation is likely to be an arduous task, requiring a multitude of methods and numerous studies. In the end, it is likely to be the overlap between results and cross validation of methods that will determine the most profitable approaches. It seems to the writer that several strategies are suitable for addressing this question, most of which already have been applied in some form. First is the use of traditional classificatory systems, such as the DSM and ICD. These would need to be modified to deal with individuals who lack speech and those who have developed from a very different "subculture" (see Introduction). However, considerable sensitivity and insight into the impact of mental retardation on a person probably would be needed to advance the application and utility of this approach. Second, it is very possible that multivariate studies, used on a much larger scale than employed thus far, may uncover syndromes and disorders whose composition and expression heretofore have not been appreciated. In order to have some likelihood of success, however, such an approach would have to include large subgroups of persons having the given disorders, and appropriate symptomatology would need to be encompassed within the descriptive content that was analyzed. This would require considerable thought and creativity before launching such an investigation. However, the payoff, in terms of understanding the nature and structure of psychopathology in mentally retarded persons, is likely to be substantial.

A third approach might take advantage of biochemical markers that have been found to have some utility in adult psychiatry. Examples are the Dexamethasone Suppression Test (DST) (Carroll et al., 1981), and serum thyrotropin response to thyrotropin-releasing hormone (Loosen & Prange, 1982). There is a variety of problems in the use of these measures, even in the assessment of nonretarded patients (regarding the DST, see APA Task Force, 1987; Arana, Baldessarini, & Ornsteen, 1985; Kraus, Grof, & Brown, 1988; for serum thyrotropin response, see Loosen & Prange, 1982). Nevertheless, they may provide very important insights in this field, especially in dealing with nonverbal individuals. At least two studies thus far have reported the use of the DST to investigate depression in intellectually handicapped patients (Pirodsky et al., 1985; Sireling, 1986). It is most interesting that traditional diagnostic criteria were unreliable in detecting positive DST responders, who presented with quite different symptom patterns than those found for depression in the classical diagnostic systems. Indeed, the dominant pattern for DST nonsuppressors was the existence of unprovoked aggressive/assaultive behavior, self-injurious behavior, and severely withdrawn behavior (Pirodsky et al., 1985). This led the authors to suggest that current diagnostic criteria for depression need to be revised for mentally retarded persons. It seems to the writer that leads such as this need to be pursued vigorously.

A fourth strategy is to use a family history approach to diagnosis. The idea would be that, whereas a variety of disorders (e.g., depression) tend to run in families, the *expression* of a given disorder may be altered by the presence of mental retardation. The family history method of data collection is both reliable and valid, and it is less expensive than the family study method of data collection (Andreason, Rice, Endicott, Reich, & Coryell, 1986). Thus, by studying a large group of persons with mental retardation, some of whom have an extensive family susceptibility for a given disorder, it may be possible to isolate important clinical markers for that disorder associated with mental retardation. Finally, a further possible strategy calls for the application of two or more of the preceding approaches in unison. We can conceptualize this as an enrichment process in which the likelihood that an individual will have a given disorder tends to increase each time a new criterion is added.[1] Thus, a person with an extensive family history of depressive disorders is more likely to be depressed if he or she is also found to have positive DST results. Such persons could be compared systematically with individuals who fulfill neither of the criteria to see if a useful symptom complex becomes evident.

Thus, it would appear that a number of viable approaches do exist for the study of behavioral and emotional problems in this population. In the end, the most fruitful tactic will be revealed empirically – that is, the results will justify the methods adopted. Nevertheless, it seems to the writer that classic diagnostic approaches and their modifications have not taken us very far in our understanding of mental disorders in the more severe forms of mental retardation. With this in mind, it is probably time to pursue the remaining strategies more aggressively.

Recommended Instruments

As a final exercise, we shall attempt to make recommendations for the most suitable scales for screening and diagnostic purposes. This process will be aided by reference to Tables 3 through 7, where the available instruments are broken down by salient features. These features of necessity will be very influential in determining an investigator's choice of instruments for any given study. The selection process should be mediated in part by the age of the population to be studied, the level(s) of mental retardation represented, types of raters available, and the degree of elaboration (diagnostic detail) desired. In the discussion to follow, the special purpose instruments have been excluded from consideration. Obviously, these merit serious consideration when assessment for a specific disorder or behavior problem is being contemplated, but they were not intended to be used for general, wide-ranging classification purposes. The several adaptive behavior scales also have been excluded because their value for screening and diagnostic ends generally is unresearched.

Instruments for assessing children. The classification of instruments by age group is shown in Table 4. When special purpose tools and adaptive behavior scales are excluded from consideration, only eight instruments remain which are suitable for assessing children. The following potential problems exist for these instruments.

---

[1] The author would like to acknowledge John Gale, M.D., (Fairview Training Center, Salem, Oregon) as the source of this idea, which emerged from a number of discussions regarding strategies for studying psychopathology in mental retardation.

(1) The ABC – Relatively small norm groups for children; relevance to noninstitutional settings is uncertain. (The ABC is being validated and normed with noninstitutionalized mentally retarded children at the time of this writing, but the outcome is not yet known.)

(2) BSAB-II – Labor intensive (requiring direct observation and extensive training of personnel); developed solely with severely/profoundly retarded individuals; lack of data outside of institutions.

(3) DCBRS – Developed in mid 1960s, probably now out of date; structure overly refined; small subscale sizes probably unacceptable to achieve adequate reliability.

(4) DDCBCL – Although the research plan for developing this instrument sounds promising, scale not available at time of this writing.

(5) EDRS-DD – Emphasizes affective/mood disorders to the exclusion of others; no standardization data; relatively few psychometric data.

(6) Fairview – Psychometric characteristics unknown; no standardization data; very lengthy.

(7) HBS Schedule – A general lack of normative or standardization data; instrument largely focused on symptomatology associated with autism; relatively few psychometric data available.

(8) PBQ – Developed on a nonretarded population; confined to preschool children; number of dimensions assessed very small.

Thus, it appears that it would be premature at this time to recommend one or more instruments for general assessment purposes in children with mental retardation. It is clear that the refinement of existing tools or the development of new ones should be a high research priority. It is possible that the appearance of the DDCBCL or the refinement of the ABC will help to rectify this situation in the near future.

Instruments for assessing adolescents and adults. Reference to Table 4 indicates that there are approximately 30 instruments suitable for assessing adolescents and adults. When special purpose tools and adaptive behavior scales are dropped, 15 instruments remain within the adolescent and adult groups. Instruments regarded by the writer as most suitable are presented in terms of the objective of assessment.

For *screening* purposes, the Reiss Screen for Maladaptive Behavior is clearly the front-runner. Indeed, it is the only instrument developed and promoted exclusively for this purpose, and its psychometric characteristics are relatively robust. However, this scale is not intended to render a specific diagnosis. Also, the relevance of DSM-III derived categories to the full range of mental retardation is poorly understood.

In terms of rating scales for *broad dimensions of behavior*, the Aberrant Behavior Checklist, the Behaviour Disturbance Scale, and the Strohmer-Prout Behavior Rating Scale are recommended. The former two are probably best suited to moderately and severely retarded individuals, whereas the SPBRS was developed for individuals with borderline IQs or mild retardation. If the concern is with affective or mood disorders, the Emotional Disturbance Rating Scale perhaps may be considered, also.

Informants often are found to be relatively insensitive to and unreliable for assessing internalizing problems such as high anxiety, tension, and depression (Costello, 1990; Shaffer et al., 1988). For this reason, it is often desirable to obtain self ratings from the person being assessed. The Prout-Strohmer Personality Inventory is one of the few self-rating instruments available in this field, but it appears to have reasonably good psychometric characteristics. If the person being evaluated is an adolescent, the Adolescent Behavior Checklist also might be considered.

Finally, for the establishment of *classical categorical diagnoses*, the following instruments might be considered: (1) Clinical Interview Schedule, (2) the Diagnostic Assessment for the Severely Handicapped (DASH) Scale, (3) the Gilson-Levitas Criteria, (4) the Psychopathology Instrument for Mentally Retarded Adults, and (5) the Structured Clinical Interview. However, *it should be noted that the psychometric characteristics for the more established instruments have* ***not*** *been impressive,* and the newer instruments (the Gilson-Levitas, the DASH, and the SCI) *are largely unstudied at this time*. Therefore, it is with great reluctance that any recommendation is made at this time. The use of such instruments for making specific diagnoses is probably most suitable for mildly and (possibly) moderately retarded individuals at the present. Their application with more severely retarded individuals assumes a level of knowledge about the structure and expression of behavioral and mental disorders that we simply do not possess at this time.

It should be noted that these suggestions are based upon the available data and instruments at the time of this review. Future refinements of these tools or the emergence of new instruments may alter the situation markedly.

Recommendations for Future Research

It is clear from the foregoing that there is a relative shortage of instruments for the assessment of behavior disorders in children with mental retardation. There also is a need to refine and extend the standardization of the most promising existing tools. Finally, and most importantly, there is a serious need to study the very nature of psychopathology across the full spectrum of mental retardation and with provision for capturing conditions that are perhaps less prevalent. These are all important goals that researchers in the field should seriously consider addressing and to which national funding bodies may wish to give special emphasis in the future.

References

Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin, 85*, 1275-1301.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross informant correlation for situational specificity. *Psychological Bulletin, 101*, 213-232.

Aman, M. G., Singh, N. N., Stewart, A. W., & Field, C. J. (1985). The Aberrant Behavior Checklist: A behavior rating scale for the assessment of treatment effects. *American Journal of Mental Deficiency, 89*, 485-491.

APA Task Force on Laboratory Tests in Psychiatry (1987). The dexamethasone suppression test: An overview of its current status in psychiatry. *American Journal of Psychiatry, 144*, 1253-1262.

Arana, G. W., Baldessarini, R. J., & Ornsteen, M. (1985). The dexamethasone suppression test for diagnosis and prognosis in psychiatry. *Archives of General Psychiatry, 42*, 1193-1204.

Andreason, N. C., Rice, J., Endicott, J., Reich, T., & Coryell, W. (1986). The family history approach to diagnosis. How useful is it? *Archives of General Psychiatry, 43*, 421-429.

Balthazar, E. E., & English, G. E. (1969). A factorial study of unstructured ward behavior. *American Journal of Mental Deficiency, 74*, 353-360.

Behar, L. B., & Stringfield, S. A. (1974). A behavior rating scale for the preschool child. *Developmental Psychology, 10*, 601-610.

Carroll, B. J., Feinberg, M., Greden, J. F., Tarika, J., Albala, A. A., Haskett, R. F., James, N. M., Kronfol, Z., Lohr, N., Steiner, M., de Vigne, J. P., & Young, E. (1981). A specific laboratory test for the diagnosis of melancholia: Standardization, validation, and clinical utility. *Archives of General Psychiatry, 38*, 15-22.

Cicchetti, D. V., & Sparrow, S. S. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*, 127-137.

Costello, E. J. (1990). Child psychiatric epidemiology: Implications for clinical research and practice. In B. Lahey & A. Kazdin (Eds.), *Advances in clinical child psychology* (Vol. 13, pp. 53-90). Greenwich, CT: JAI Press.

Espie, C. A., Montgomery, J. M., & Gillies, J. B. (1988). The development of a psychosocial behaviour scale for the assessment of mentally handicapped people. *Journal of Mental Deficiency Research, 32*, 395-403.

Kleinbaum, D. G., Kupper, L. L., & Morgenstern, H. (1982). *Epidemiologic research. Principles and quantitative methods.* London: Lifetime Learning Publications.

Krauss, R. P., Grof, P., & Brown, G. M. (1988). Drugs and the DST: Need for a reappraisal. *American Journal of Psychiatry, 145*, 666-674.

Leudar, I., Fraser, W. I., & Jeeves, M. A. (1984). Behaviour disturbance and mental handicap: Typology and longitudinal trends. *Psychological Medicine, 14*, 923-935.

Loosen, P. T., & Prange, A. J. (1982). Serum thyrotropin response to thyrotropin-releasing hormone in psychiatric patients: A review. *American Journal of Psychiatry, 139*, 405-416.

Matson, J. L., Coe, D. A., Gardner, W., & Sovner, R. (1990). *A factor analytic study of the Diagnostic Assessment for the Severely Handicapped (DASH) scale.* Unpublished manuscript, Louisiana State University, Baton Rouge.

Nihira, K. (1978, May). *Dimensions of maladaptive behavior in institutionalized mentally retarded persons.* Paper presented at the Fourth Annual Western Conference on Mental Retardation, "The Profoundly Mentally Retarded," Corpus Christi, TX.

Pirodsky, D. M., Gibbs, J. W., Hesse, R. A., Hsieh, M. C., Krause, R. B., & Rodriguez, W. H. (1985). Use of the dexamethasone suppression test to detect depressive disorders of mentally retarded individuals. *American Journal of Mental Deficiency, 90*, 245-252.

Prout, H. T., & Strohmer, D. C. (1989). *Prout-Strohmer Personality Inventory manual.* Schenectady, NY: Genium Publishing Corporation.

Quay, H. C. (1979). Classification. In H. C. Quay & J. S. Werry (Eds.), *Psychopathological disorders of childhood* (2nd ed., pp. 1-42). New York: Wiley.

Reiss, S. (1988). *Test manual for the Reiss Screen for Maladaptive Behavior.* Orland Park, IL: International Diagnostic Systems.

Rojahn, J. (1986). Self-injurious and stereotypic behavior of non-institutionalized mentally retarded people: Prevalence and classification. *American Journal of Mental Deficiency, 91*, 268-276.

Shaffer, D., Schwab-Stone, M., Fischer, P., Davies, M., Piacentini, J., & Gioia, P. (1988). *Revised version of the Diagnostic Interview Schedule for Children: Results of a field trial and proposals for a new instrument DISC 2.* Unpublished report to the National Institute of Mental Health Division of Epidemiology, Columbia University, New York.

Sireling, L. (1986). Depression in mentally handicapped patients: Diagnostic and neuroendocrine evaluation. *British Journal of Psychiatry, 149*, 274-278.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Interview edition, expanded form manual. Vineland Adaptive Behavior Scales.* Circle Pines, MN: American Guidance Service.

Spivack, G., Haimes, P. E., & Spotts, J. (1967). *Devereux Adolescent Behavior Rating Scale manual.* Devon, PA: The Devereux Foundation.

Spivack, G., & Spotts, J. (1966). *Devereux Child Behavior Rating Scale manual.* Devon, PA: The Devereux Foundation Press.

Strohmer, D. C., & Prout, H. T. (1989). *Strohmer-Prout Behavior Rating Scale manual.* Schenectady, NY: Genium Publishing Corporation.

Tables

Table 1

Magnitude of Interrater Reliability Correlations (Using Cicchetti & Sparrow Criteria) for Different Types of Informant

	Similar Informants	Different Types of Informants	Subjects' Self-Ratings
<.40 (Poor)	10 (15%)	127 (65%)	2 (29%)
.40-.59 (Fair)	22 (33%)	55 (28%)	1 (14%)
.60-.74 (Good)	23 (35%)	8 (4%)	0 (0%)
≥.75 (Excellent)	11 (27%)	6 (3%)	4 (57%)

Note. Figures are abstracted from a meta-analysis conducted by Achenbach, McConaughy, and Howell (1987) . Self Ratings differ from the remainder, as they reflect test-retest comparisons rather than interrater comparisons. Self ratings also may be included under the different informant column but only if self ratings were paired with a third-party rating. Figures in parentheses represent percentages of column totals.

Table 2.

Psychometric Features Addressed in Reviewed Instruments (for Part I only)

	Mentally Retarded Subjects in Standardization Group?				Reliability				Validity		
Instrument	Mild	Moderate	Severe	Prof.	Alpha/ Split-Half	Item-Total	Test-Retest	Inter-Rater	Factorial/ taxonomic	Criterion Group	Congruent
AAMD Adaptive Behavior Scale: Residential and Community Edition	X	X	X	X	X	X	X	X		X	X
AAMD Adaptive Behavior Scale: School Edition	X	X	X	?	X				X	X	?
Aberrant Behavior Checklist	X	X	X	X	X	X	X	X	X	X	X
Adolescent Behavior Checklist	X				X		X	NA	X	X	X
Balthazar Scales of Social Adaptation			X	X				X	X		
Behaviour Disturbance Scale	X	X	X				X	X	X	X	X
Client Development Evaluation Report	X	X	X	X				X	X	X	X
Clinical Interview Schedule	X	X	X	X			X	X	X	X	X
Devereux Adolescent Behavior Rating Scale	?	?	?	?	X		X	X	X	X	
Devereux Child Behavior Rating Scale	X	X	X	X			X	X	X	X	

Instrument	Mentally Retarded Subjects in Standardization Group?				Reliability				Validity		
	Mild	Moderate	Severe	Prof.	Alpha/ Split-Half	Item-Total	Test-Retest	Inter-Rater	Factorial/ taxonomic	Criterion Group	Congruent
Diagnostic Assessment for the Severely Handicapped			X	X	X			X	X		
Emotional Disorders Rating Scale			X	X	X		X	X	X	X	X
Minnesota Developmental Programming System: Behavior Management Assessment	X	X	X	X				X			X
Preschool Behavior Questionnaire	?	?	?	?					X		X
Prout-Strohmer Personality Inventory	X				X	X	X	NA	X	X	X
Psychopathology Instrument for Mentally Retarded Adults	X	X	X	X	X	X	X	X	X	X	X
Reiss Screen	X	X	X	X	X			X	X	X	X
Schedule of Handicaps, Behaviour, & Skills		X	X	X				X		?	
Self-Report Depression Questionnaire	X	X	X		X	X	X	NA	X		X
Strohmer-Prout Behavior Rating Scale	X				X	X		X	X	X	X
Vineland Adaptive Behavior Scales	X	X	X	X	X		X	X		X	

Table 3

Instruments in Parts I and II Classified by Type of Rater

Rater Type	Instruments	Number
Self	Adolesc. Behav. CL, PSPI, PIMRA, *RCMAS*, *SJS*, *SRDQ*, *Zung*	7
Informant	ABS:R, ABS:S, ABC, *Attention CL*, BDS$_1$, BDS$_2$, BeERS, BIRD, *BPI*, CDER, *Comm. Style Q*, DABRS, DASH, DCBRS, DDCBCL, EDRS-DD, Fairview, Gilson-Levitas, HBS, *MAS*, MDPS-BMA, PBQ, PBS, PIMRA, Reiss, *Soc.Part. RS*, SPBRS, Vineland	28
Diagnoser	BSAB-II[a], CIS, *SAP*, SCI, SEBI [b]	5

Note. Scale abbreviations are defined fully in Appendix C. Italicized instruments are special purpose scales.

[a] Trained observers required.

[b] Trained personnel extract relevant information from existing written records.

Table 4

Instruments in Parts I and II Classified by Age Group Covered

Age Group	Instruments	Number
Children[a]	ABS:R, ABS:S, ABC, *Attention CL*, BIRD[b], BSAB-II, BDS$_2$, *BPI*, CDER, DCBRS, EDRS-DD, Fairview, HBS, *MAS*, MDPS-BMA, PBQ, *RCMAS*, SEBI, Vineland	19
Adolescents	ABS:R, ABS:S, ABC, Adoles. Behav. CL, BDS$_1$, BDS$_2$, BeERS, BIRD[b], BPI, CDER, *Comm. Style Q*, DABRS, DASH, DDCBCL, EDRS-DD, Fairview, HBS, *MAS*, MDPS-BMA, PIMRA, PSPI, Reiss, SEBI, SPBRS, *SRDQ*, Vineland	26
Adults	ABS:R, ABC, BDS$_1$, BDS$_2$, BeERS, BIRD[b], *BPI*, CDER, CIS, *Comm. Style Q*, DASH, Fairview, Gilson-Levitas, HBS, *MAS*, MDPS-BMA, PBS, PIMRA, PSPI, Reiss, *SAP*, SCI, SEBI, *SJS*, *Soc. Part. RS*, SPBRS, *SRDQ*, Vineland, *Zung*	29

Note. Scale abbreviations are defined fully in Appendix C. Italicized instruments are special purpose scales.

[a] Excluding the DDCBCL, which is not yet fully developed at the time of this writing.

[b] Due to its relationship to the Vineland, the BIRD is treated as an adaptive bahvior scale in the relevant discussions.

Table 5

Instruments in Parts I and II Classified by Method of Derivation

Method of Derivation	Instruments	Number
Empirical	ABC, BDS$_1$, BDS$_2$, BSAB-II *Comm. Style Q*, DABRS, DCBRS, DDCBCL[a], PBQ, PBS	10
Clinical/*A Priori*	ABS:R, ABS:S, Adolesc. Behav. CL, *Attention CL*, BeERS, CIS, EDRS-DD, Fairview,Gilson-Levitas, HBS, *MAS*, MDPS-BMA,PIMRA, *SAP*, SCI, *Soc. Part. RS*, SEBI, *SRDQ*, *Zung*	19
Clinical & Empirical	BIRD, *BPI*, DASH, PSPI, *RCMAS*, Reiss, SPBRS	7

Note. Scale abbreviations are defined fully in Appendix C. Italicized instruments are special purpose scales.

[a] Not available at the time of this writing, although the research plan calls for derivation of subscales by factor analysis.

Table 6

Instruments in Parts I and II Classified by Number of Subscales

Number of Subscales	Instruments	Number
1	*Attention CL*, BeERS, SEBI, *Soc. Part. RS*, *SRDQ*, *Zung*	6
2	BDS$_2$, BIRD, Vineland	3
3	*BPI*, PBQ, *RCMAS*	3
4	*MAS*, *SJS*	2
5 & 6	ABC, BDS$_1$, DASH[a], Fairview, PBS, PSPI	6
7 & 8	Adolesc. Behav. CL, BSAB-II, EDRS-DD, Gilson-Levitas, PIMRA, Reiss, SCI	7
9 & 10	CIS, *SAP*	2
11-15	ABS:R, ABS:S, *Comm. Style Q*, DABRS, DASH[a], HBS, SPBRS	7
17	DCBRS	1

Note. Scale abbreviations are defined fully in Appendix C. Italicized instruments are special purpose scales. For instruments having both adaptive and maladaptive sections, only the numbers of maladaptive subscales are summarized here.

MDPS-BMA not included because of uncertainty regarding number of dimensions sampled.

[a] As DASH can be scored according to factorial or *a priori* scoring methods, it appears twice here.

Table 7

Instruments in Parts I and II Classified by Level of Mental Retardation Covered

Level of Retardation	Instruments	Number
Mild	ABS:R, ABS:S, ABC (?), Adolesc. Behav. CL, *Attention CL*, BDS$_1$, BDS$_2$, *BPI*, *Comm. Style Q*, CIS, DABRS, DCBRS, EDRS-DD, Fairview, Gilson-Levitas, HBS, *MAS*, MDPS-BMA, PBS, PIMRA, PSPI, *RCMAS*, Reiss, *SAP*, SCI, *SJS*, SPBRS, *SRDQ*, Vineland, *Zung*	30
Moderate	ABS:R, ABS:S, ABC, BDS$_1$, BDS$_2$, *BPI*, CIS, *Comm. Style Q*, DABRS (?), DCBRS, EDRS-DD, Gilson-Levitas, HBS, *MAS*, MDPS-BMA, PBS, PIMRA, *RCMAS (?)*, Reiss, *SAP*, *SJS*, *SRDQ (?)*, Vineland, *Zung*	24
Severe	ABS:R, ABS:S, ABC, BDS$_1$, BDS$_2$, *BPI*, BSAB-II, CIS, *Comm. Style Q*, DASH, DCBRS, EDRS-DD, Fairview, HBS, *MAS*, MDPS-BMA, PIMRA, Reiss, Vineland	19
Profound	ABS:R, ABS:S(?), ABC, BDS$_2$, *BPI*, BSAB-II, CIS, DASH, DCBRS(?), EDRS-DD, Fairview, HBS, *MAS*, MDPS-BMA, PIMRA(?), Reiss, Vineland	17

Note. Scale abbreviations are defined fully in Appendix C. Italicized instruments are special purpose scales. (?) signifies uncertainty regarding suitability of the instrument for this population. BeERS, BIRD, and SEBI not included because of uncertainty concerning relevant target groups.

Table 8

Commonalities among Factors from Factor Analytic Research

General Description	Related Factors/Instruments
1. Aggressive, Antisocial, Self-Injurious	ABS:R: Temper tantrums Violent and antisocial Destructive behavior (including SIB[a]) Rebellious behavior
	ABC: Irritability, agitation, crying (including SIB[a])
	BSAB-II: Responds aggressively to staff/peers
	DASH: Antisocial Emotional lability (includes screaming, crying, and hair pulling)
	PBQ: Hostile-aggressive
	BDS_1: Agressive conduct, Antisocial conduct
2. Social Withdrawal	ABS:R: Withdrawal
	ABC: Lethargy, social withdrawal
	BSAB-II: Failure to respond
	BDS_1: Mood disturbance
	DASH: Social withdrawal (includes repetitive movements)
	PBQ: Nil
3. Stereotypic Behavior	ABS:R: Stereotyped and hyperactive behavior
	ABC: Stereotypic behavior
	BDS_1: Idiosyncratic mannerisms
	BSAB-II: Stereotypy, posturing behaviors
	DASH: Social withdrawal (includes repetitive movements, activities, and sounds)
	PBQ: Nil

4. Hyperactivity	ABS:R: Stereotyped and hyperactive behavior
	ABC: Hyperactivity, noncompliance
	BSAB-II: Disorderly, nonsocial behavior
	BDS_1: Communicativeness[b]
	DASH: Nil
	PBQ: Hyperactive-distractible
5. Repetitive verbalization	ABS:R: Nil
	ABC: Inappropriate speech
	BDS: Nil
	BSAB-II: Isolated repetitious verbalization
	DASH: Language disorder
	PBQ: Nil
6. Anxious, tense, and fearful	ABS:R: Withdrawal
	ABC: Lethargy, social withdrawal
	BDS_1: Nil
	DASH: Emotional lability (includes crying, screaming, mood changes) Sleep disorder
	PBQ: Anxious, fearful
7. Self-Injurious Behavior	ABS:R: Destructive behavior (including SIB[a])
	ABC: Irritability, agitation, crying (including SIB[a])
	BDS_1: Self injury
	DASH: Emotional lability (includes crying, screaming, hair pulling)
	PBQ: Nil

Note. Full scale names are as follows: ABS:R = AAMD Adaptive Behavior Scale: Residential and Community Edition (Nihira, 1978); ABC = Aberrant Behavior Checklist (Aman, Singh, Stewart, & Field, 1985); BDS_1 = Behaviour Disturbance Scale (Leudar, Fraser, & Jeeves, 1984); DASH = Diagnostic Assessment for the Severely Handicapped (Matson, Gardner, Coe, & Sovner, 1990); PBQ = Preschool Behavior Questionnaire (Behar & Stringfield, 1974).

[a] SIB = Self-Injurious Behavior

[b] Communicativeness factor includes: "Is inattentive," "Is overactive," "Does not take part in group activities."

Appendices

Appendix A

Societies and Associations Whose Memberships Were Notified Regarding the Review

1. The American Academy of Child and Adolescent Psychiatry (via Academy Newsletter).

2. Area 25 (Applied Behavior Analysis) of the American Psychological Association.

3. Area 33 (Psychology of Mental Retardation) of the American Psychological Association.

4. Sections 1 and 5 (Clinical Child Psychology, and Pediatric Psychology, respectively) of Area 12 (Clinical Psychology) of the American Psychological Association.

5. The Association for the Advancement of Behavior Therapy (Notice appeared in the *Behavior Therapist* in June, 1989).

6. Australian Society for the Study of Intellectual Disability (via Society newsletter).

7. Behavior Pediatrics Society (Notice appeared in the *Journal of Developmental and Behavioral Pediatrics).*

8. The British Institute of Mental Health (via Institute Newsletter).

9. All Mental Retardation Research Centers and University Affiliated Programs for Persons with Developmental Disabilities (Electronic mail notice, sent on 13 Feb. 1989).

10. Society for Research in Child and Adolescent Psychopathology (selected members only, as the Society had no newsletter at the time of the review).

Appendix B: Summary of Psychometric Characteristics of Reviewed Scales

Instrument	Authors	Samples	RELIABILITY: Internal Consistency (alpha)	Item Total Correlations
AAMD Adaptive Behavior Scale: Residential and Community Edition (ABS:R)	1. Nihira, Foster, Shellhaas, & Leland, 1975 Also Meyers, Nihira, Zetlin, 1979	a) 4,014 institutional residents, ranging in age from 3 to 69 years. b) 133 residents of three training schools. c) 919 adults and 313 children and adolescents residing in state institutions.		
	2. Nihira, 1978	2,616 institutional residents, aged 10 to 69 years, with mild through profound mental retardation.		
	3. Isett & Spreat, 1979	Groups of 28 and 29 institutional residents.		
	4. Spreat, 1980	Formerly institutionalized subjects (N=95), institutional residents referred for discharge (N=97), and 178 current residents.		

Note. All citations appearing in this section are referenced in their respective sections in Part I.

		VALIDITY:		
Test-Retest	Interrater	Factorial/ Taxonomic	Criterion Group	Congruent
	Sample b: Phi coefficients ranged from .37 (Unacceptable Vocal Habits) to .69 (Untrustworthy Behavior) across subscales. Mean=.55.	Sample c: Parts I & II, together, found to be measuring three orthogonal dimensions: Personal Independence, Social Maladaption, and Personal Maladaption.	Some Part II domain scores discriminated among subjects in five different administrative placement units.	
		Nine meaningful factors derived as follows: (1) Violent and Antisocial Behavior, (2) Rebellious Behavior (3) Untrustworthy Behavior, (4) Destructive Toward Property and Self, (5) Stereotyped and Hyperactive, (6) Inappropriate Body Exposure, (7) Withdrawal, (8) Inappropriate Sexual Behavior, and (9) Temper Tantrums.		
2-week reliability (Spearman correlations) ranged from .60 (Inappropriate Mannerisms) to .97 (Withdrawal). Mean r_S = .83.	Spearman correlations ranged from .32 (Untrustworthy Behavior) to .84 (Stereotyped and Odd Mannerisms). Mean=.56.			
				Parts I and II used to predict current placement. Seven domains (including Untrustworthy Behavior, Unacceptable Vocal Habits, Psychological Disturbances) predicted 60% of derivation sample and 49% of cross validation sample. Using factorially derived scores, 54% of both derivation and cross validation samples correctly classified.

Instrument	Authors	Samples	RELIABILITY: Internal Consistency (alpha)	Item Total Correlations
(ABS:R cont.)	5. Clements, DuBois, Bost, & Bryan, 1981	210 institutional residents rated by seven psychologists.		
	6. Bean & Roszkowski, 1982	265 institutional residents, aged 7 to 53 years, and with mild to profound mental retardation.	Alpha ranged from .64 (Self-abusive Behavior) to .92 (Antisocial Behavior) with a mean of .78.	45 items correlated < .30 with their own domain; 40 items correlated higher with other domains, and 35% of items were regarded as possessing undesirable characteristics (excluding Medication domain). 62% of Part II items judged satisfactory.
	7. Salagaras & Nettelbeck, 1983	550 students, aged 13 to 20 years, attending Special Schools in South Australia.		
	8. Stack, 1984	90 adults, aged 18 to 51 years, with mild through profound mental retardation, and living in various types of residential settings. Informants worked with the subjects in parallel roles and at similar times of the day.		

Test-Retest	**Interrater**	**VALIDITY: Factorial/ Taxonomic**	**Criterion Group**	**Congruent**
				Global ratings of severity of mental disturbance correlated weakly with Part II frequency ratings (r=.43). Correction using severity weightings increased mean correlation to .54.
	Phi coefficients ranged from .36 (Unacceptable Vocal Habits) to .61 (Rebellious Behavior; Sexually Aberrant Behavior) with a mean of .49 across domains.		Subjects with Down sydrome rated significantly lower on Hyperactive Tendencies domain. Subjects residing in institutional settings received higher ratings on Violent & Destructive Behavior, Antisocial Behavior, and Sexually Aberrant Behavior. Subjects receiving medication were rated higher on Violent & Destructive Behavior, Unacceptable/Eccentric Habits, and Psychological Disturbances.	
	Single-score intraclass correlations ranged from .25 (Hyperactive) to .70 (Violent & Destructive). Mean correlation = .51.			

Instrument	Authors	Samples	RELIABILITY: Internal Consistency (alpha)	Item Total Correlations
(ABS:R cont.)	9. Aman, Singh, Stewart, & Field, 1985	70 institutional residents with moderate through profound mental retardation.		
AAMD Adaptive Behavior Scale–School Edition (ABS:S)	1. Lambert & Nicoll, 1976.	Children aged 7 to 13 years with one of four educational placements as follows: Regular classes (N=1,157), Educable Mentally Retarded classes (N=880), Trainable Mentally retarded classes (N=185), and Educationally Handicapped classes (N=396).		
	2. Lambert, 1981.	6,500 children, aged 3 to 16 years, placed in Regular, Educable Mentally Retarded, and Trainable Mentally Retarded classes.	Alpha coefficients for factor 4 (Social Adjustment) ranged from .77 to .97 across age and sex (median alpha in .90s). For factor 5 (Personal Adjustment) alpha ranged from .27 to .80 with most values in the .50s and .60s. No alpha coefficients presented for domain scores.	
	3. Lambert & Hartsough, 1981.	Children aged 7 to 17 years having one of three educational placements, as follows: Regular classes (N=1,650), Educable Mentally Retarded classes (N=3,052), and Trainable Mentally Retarded classes (N=828).		

Test-Retest	Interrater	VALIDITY: Factorial/ Taxonomic	Criterion Group	Congruent
				Comparison of ratings on ABS and Aberrant Behavior Checklist resulted in the following correspondences: Self Abusive Behavior and Irritability (r_S=.59); Withdrawal and Lethargy/Withdrawal (.69); Stereotyped Behavior and Stereotypic Behavior (.69); Unacceptable Vocal Habits and Inappropriate Speech (.42). No correspondence between Hyperactive Tendencies and Hyperactivity.
		Factor analysis of domain scores across school classification and across age groups produced the same four factors, as follows: (1) Functional Autonomy, (2) Interpersonal Adjustment, (3) Social Responsibility, and (4) Intrapersonal Adjustment.		
		Several factor analyses rendered three adaptive and two maladaptive factors, respectively, as follows: (1) Personal Self-sufficiency, (2) Community Self-sufficiency, (3) Personal-Social Responsibility, (4) Personal Adjustment, and (5) Social Adjustment.		Factors based on Part II domains had low-to-moderate correlations with achievement measures.
			Using a discriminant analysis, a composite score was calculated for predicting school classification (Regular, EMR, TMR). Using a cross validation procedure, between 63% and 79% (median=74%) of children were correctly classified on the basis of ABS:SE scores.	

Instrument	Authors	Samples	RELIABILITY: Internal Consistency (alpha)	Item Total Correlations
Aberrant Behavior Checklist (ABC)	1. Aman, Singh, Stewart, & Field, 1985a	927 ambulatory residents in institutions, having moderate, severe, and profound mental retardation. Sample constituted about 1/3 of residential population of New Zealand. Average subscale scores and SDs presented for sample.		
	2. Aman, Singh, Stewart, & Field, 1985b	Same as #1, above.	Alpha coefficients across five subscales ranged from .86 to .94 (M=.91).	
	3. Aman, Richmond, Stewart, Bell, & Kissel, 1987	937 residents of New Zealand institutions and 531 residents of a U.S. developmental center. Subjects had moderate through profound retardation. U.S. sample constituted 82% of institution's ambulatory population.	For U.S. sample, alpha coefficients ranged from .88 to .94 (M=.90).	
	4. Aman, Singh, & Turbott, 1987	28 subjects in each of four residential units (N=112), with moderate through profound mental retardation.		

Test-Retest	Interrater	VALIDITY: Factorial/ Taxonomic	Criterion Group	Congruent
		Two independent factor analyses yielded the same set of 5 subscales. The analyses accounted for 71% and 76% of the common variance.		
Reportedly very high. Later discounted for methodological reasons (Aman, Singh, & Turbott, 1987).	Nine rater pairs rated 25 subjects each. Correlations ranged from .17 to .90. Mean correlation across all raters and subscales = .63.		Subjects attending training facilities had lower scores than non-attenders on all subscales except Inappropriate Speech. Subjects with Down syndrome had significantly lower scores on all except Lethargy, Social Withdrawal subscale. Higher scores on all except one subscale were associated with taking psychoactive drugs.	ABC subscale scores correlated in predictable ways with Fairview Self-Help Scale, Vineland Social Maturity Scale, and AAMD Adaptive Behavior Scale. ABC scores not correlated with IQ. All except one subscale score were correlated with direct observations of component behaviors.
		Factor structure of ratings for U.S. sample nearly identical to the original factor solution. Coefficient of congruence ranged from .88 to .96 (M= .93). 50 of 58 items loaded with same respective factors as in original analysis.	Subjects with epilepsy rated as more disturbed on Irritability and Hyperactivity subscales. Subjects with a diagnosis of psychosis had higher scores on all subscales. Psychoactive drug use associated with higher scores on all subscales.	
12 nurses/nurse-aides rated 28 residents each at 4-weekly intervals. Across all modes of instruction, correlations ranged from .71 to .81 for all subscales (M= .77).	Across groups of raters, type of instruction, and time of rating, correlations ranged from .23 to to .97 (M=.58).			

Instrument	Authors	Samples	RELIABILITY:	
			Internal Consistency (alpha)	Item Total Correlations
(ABC cont.)	5. Newton & Sturmey, 1988	209 adults residing in residential institutions in England. 45% were nonambulatory. Level of mental retardation not reported.	Alpha coefficients ranged from .84 to .92 for the 5 subscales (M=.89).	Item whole correlations ranged from .39 to .88 (M=.60).
	6. Raft & Richmond, 1989	32 profoundly retarded institutionalized residents.		
	7. Freund & Reiss, 1990	110 children, adolescents, and young adults, with borderline IQ to severe mental retardation, attending a neuropsychiatric unit.	Parent ratings: alpha coefficients ranged from .83 to .93 (M=.89). Teacher ratings: Alpha coefficients ranged from .79 to .94 (M=.88) for the 5 subscales.	
	8. Sturmey & Ley, 1990	24 mentally retarded adults attending a clinic for psychiatric, behavioral, and/or medical problems.		

Test-Retest	Interrater	VALIDITY:		
		Factorial/	Criterion Group	Congruent
		78% of items loaded on same factors as in original study (Aman et al., 1985a). When items were scored dichotomously (occurred vs. did not occur), 81% of items loaded with same factors as in original study. 55% of the variance explained by the 5 factors.		
			Significantly more subjects with positive Dexamethasone Suppression Test (DST) results had high Irritability scores than did DST suppressors.	
Parent ratings: correlations ranged from .80 to .95 (M=.89) for 30 subjects. Teacher ratings: correlations ranged from .50 to .67 (M=.60) for 25 children.	Correlations between parent and teacher ratings ranged from .18 to .49 (M=.40).	49 of 54 items analyzed (91%) from parent ratings loaded on same factors as in original study (Aman et al., 1985a). 44 of 54 items analyzed (80%) from teacher ratings loaded on same factors as in original study.		
				Several subscales from the ABC correlated with subscales on the Psychopathology Instrument for Mentally Retarded Adults as follows: Lethargy & Schizophrenic/ Affective/Somatoform disorders; Stereotypic Behavior & Personality Disorder/ Inappropriate mental adjustment; Hyperactivity & Adjustment disorder.

Instrument	Authors	Samples	RELIABILITY: Internal Consistency (alpha)	Item Total Correlations
(ABC cont.)	9. Bihm & Poindexter, in press	470 moderately to profoundly retarded residents of an ICF/MR facility.	Coefficient alpha ranged from .84 to .93 (M=.89)	
	10. Rojahn & Helsel, in press	204 mentally retarded inpatients in psychiatric unit. Age range 3 to 23 years (M=10.7); level of mental retardation ranged from borderline to profound.	Alpha coefficients ranged from .82 to .94 (M=.89) for the five subscales.	
Adolescent Behavior Checklist	1. Cosgrove-Dapuzzo, 1989	40 adolescents, aged 12 to 16 years. 20 subjects were controls, and 20 were diagnosed as emotionally disturbed and received special education services. IQs not reported, but reading levels backward by three years.	Alpha coefficients ranged from .58 (Intake/Control) to .91 (Oppositional) for subscales; mean=.76. Six of 8 subscales had alphas ≥ .70. Alpha for Lie scale =.25. Coefficient alpha for all items =.95.	

		VALIDITY:		
Test-Retest	Interrater	Factorial/ Taxonomic	Criterion Group	Congruent
		86% of items loaded most heavily on same respective factors as in the original study (Aman et al., 1985a).		
	Double ratings obtained with 56 raters on 130 subjects. Correlations ranged from .39 to .61 (M=.50).	91% of items loaded on same respective factors as in original study. 32% of variance explained by five obtained factors.	Three subscales differentiated significantly between several DSM-III diagnostic groups.	
3 week test-retest reliabilities ranged from .86 to 1.00 (M=.94) for Emotionally Disturbed group, from .93 to 1.00 (M=.99) for Normal group, and from .87 to 1.00 (M=.96) for combined groups (eight subscales and Lie scale). Change in mean rates of reported symptoms occurred for only one subscale (Affective Disorder).	N/A	Items adapted from DSM-III-R.	Emotionally Disturbed group had significantly higher scores than controls on all diagnostic groupings except Intake/ Control. No differences on Lie scale. Subscale scores not presented for subjects with specific psychiatric diagnoses.	Correspondence with Youth Self-Report Scale (YSR) (Achenbach & Edelbrock, 1987): Anxiety, Affective Illness, and Trait Disorder signif. associated with internalizing disorders; Hyperactivity, Conduct Disorder, Oppositional related to externalizing domain of YSR. Three of 8 comparisons showed signif. correlations with same respective subscales on YSR. No subscale data presented for controls. Correlation between Total score and YSR total=.90 for all subjects. Correspondence with Teacher Report Form (TRF) (Achenbach & Edelbrock, 1986): Few signif. correlations between subscales. Correlation between Total scores =.56 for all subjects.

Instrument	Authors	Samples	RELIABILITY: Internal Consistency (alpha)	Item Total Correlations
(Adolescent Behavior Checklist cont.)	2. Demb, Brier, & Huron, 1989	Data are reported for two samples: Sample 1 made up of developmentally disabled adolescents, aged 12 to 21 years, with DSM-III-R diagnoses of learning disability and/or mild mental retardation. Sample 2 made up of adolescents 16 to 21 years of age, with DSM-III-R specific developmental disorders and/or borderline intellectual functioning and arrested for non-violent offenses. Group sizes and IQs not reported.		
Balthazar Scales of Adaptive Behavior II. Scales of Social Maladaptation	1. Balthazar & English, 1969	288 severely retarded, institutional residents, aged 5 to 57 years.		
	2. Balthazar, 1973	Same as #1, above.		
Behaviour Disturbance Scale (BDS)	1. Leudar, Fraser, & Jeeves, 1984	a) BDS1: 629 subjects, aged 16 to 45 years, with mild through severe mental retardation. b) BDS2: 247 subjects, aged 15 to 52 years, with mild through severe mental retardation and residing in developmental centers (34%) or in the community (66%).		

Test-Retest	Interrater	VALIDITY: Factorial/ Taxonomic	Criterion Group	Congruent
		71 subscale items grouped by factor analysis onto 18 factors. Organization of subsequent subscales only partly determined by this solution.		
	Two studies conducted with 21 and 25 subjects. Mean "proportion agreement" ranged from .58 to .76 across subscales for study 1 and from .75 to .97 for study 2. Overall proportion agreement for maladaptive subscales was .80.			
Stability of rating scores assessed over 2 years for 118 subjects. Not all raters held constant. Original subscales or their transformations predicted between 24% and 58% of outcome variance for all except Self Injury subscale.	Derived for 10 subjects apparently across all items, r=.75. Derived for 16 subjects for each subscale and Total Score. Subscale correlations ranged from .65 to .89. Total Score r=.91.	BDS1: Six factors, five of which overlapped with BDS2, a subsequent and lengthier version of the BDS. BDS2: All six subscales factor analytically derived; factor loadings generally high (M=.57); factors accounted for 55% of variance.	Comparisons of subjects residing in hospitals (developmental centers) and those residing in the community indicated higher scores for the former on the, Aggressive Conduct, Antisocial Conduct, and Self Injury subscales.	

Instrument	Authors	Samples	RELIABILITY: Internal Consistency (alpha)	Item Total Correlations
(BDS cont.)	2. Fraser, Leudar, Gray, & Campbell, 1986	133 subjects, aged 18 to 45 years, with mild through severe mental retardation, residing in hospitals (developmental centers) (49%) and in the community (51%).		
Client Development Evaluation Report (CDER)	1. Harris, Eyman, & Mayeda, 1982.	360 subjects on the California caseload, sampled proportionately for age and level of disability.		
	2. Arias, Ito, & Tagaki, 1983	82 severely and profoundly retarded institutional residents, aged 14 to 25 years.		
	3. Widaman, Gibbs, & Geary, 1987	6,048 persons with mild through severe mental retardation and aged 1 to 83 years. Sample partitioned by type of residence, age, and level of mental retardation into 14 subgroups.		

		VALIDITY:		
Test-Retest	**Interrater**	**Factorial/ Taxonomic**	**Criterion Group**	**Congruent**
	Reliability for Total Score =.89. N not reported.		Several BDS subscales significantly associated with setting: Self Injury with hospitalization; Aggression with hospitalization.	Several BDS subscales associated with factors from Clinical Interview Schedule as follows: Communicativeness & Neurasthenia , Communicativeness & Mental Retardation (inverse correlation), and Aggression & Phobia.
	Correlations ranged between .60 and .90 for 12 of the 15 items of the Emotional Domain. Three items had correlations below .50.			
				Emotional Domain of CDER was compared with maladaptive factors on the Behavior Developmemt Survey. Positive correlation of .78 obtained.
		Factor analysis yielded six interpretable factors across samples as follows: (1) Motor Development, (2) Independent Living Skills, (3) Cognitive Competence, (5) Social (or Extra-punitive) Maladaption, and (6) Personal (or Intrapunitive) Mal-adaption. Median coefficients of congruence across samples were .96 and .95 for factors 5 and 6, respectively. Correlation between factors 5 and 6 was .72.		

Instrument	Authors	Samples	RELIABILITY: Internal Consistency (alpha)	Item Total Correlations
(CDER cont.)	4. Nihira, Price-Williams, & White, 1988	3, 975 individuals having specific dual diagnoses and 3, 975 matched controls without psychiatric disorders.		
Clinical Interview Schedule (also called Standardized Psychiatric Interview)	1. Goldberg, Cooper, Eastwood, Kedward, & Shepherd, 1970	40 hospitalized psychiatric patients without mental retardation.		
	2. Ballinger, Armstrong, Presley, & Reid, 1975	27 inpatients in a mental subnormality hospital, half able to converse and half with little or no speech. Age ranged from 15 to 70 years; 13 subjects had IQs >35 and 14 had IQs <35.		
	3. Ballinger & Reid, 1977	75 adults, mean age 28 years, with mild to severe mental retardation, attending a training center, and 75 adults, mean age 46 years and with mild through profound mental retardation, residing in a mental subnormality hospital.		

Test-Retest	Interrater	VALIDITY: Factorial/ Taxonomic	Criterion Group	Congruent
			Subjects having Adjustment Disorders rated significantly lower than controls on Social Maladaption domain. Subjects with Pervasive Developmental Disorders, Conduct Disorders Schizophrenic Disorders, and Personality Disorders rated significantly lower on both Social and Personal Maladaptation domains.	
	Pearson's r ranged from .79 to .98 (M=.85) for symptoms; from .66 to .98 (M=.89) for manifest abnormalities. Weighted kappa ranged from .67 to .81 (M=.72) for symptoms; from .48 to .94 (M=.71) for manifest abnormalities.	Developed in relation to International Classification of Diseases.		
	11 of 31 items regarded as very satisfactory, 9 as satisfactory, 6 unsatisfactory, and 6 "unproven". Correlations ranged from -.18 to .93 (M=.64) for Part 2 and from -.02 to .69 (M=.20) for Part 4.			Average agreement regarding overall severity between three raters and consultant psychiatrists =.55.
			Subjects in the hospital group had more symptoms and manifest abnormalities and significantly higher overall severity ratings than training center subjects.	

Instrument	Authors	Samples	RELIABILITY: Internal Consistency (alpha)	Item Total Correlations
(CIS cont.)	4. Reid, Ballinger, & Heather, 1978	100 institutionalized subjects with severe (n=49) and profound (n=51) mental retardation. Age ranged from 17 to 71 years (M=35).		
	5. Reid, Ballinger, Heather, & Melvin, 1984	86 adults, with severe and profound retardation, followed up after 6 years. At follow-up, age ranged from 24 to 78 years(M=41).		
	6. Fraser, Leudar, Gray, & Campbell, 1986	65 subjects from mental subnormality hospitals and 68 from community training centers. Age ranged from 18 to 65 years (M=29.5) and level of mental retardation ranged from mild to severe.		
Devereux Adolescent Behavior (DAB) Rating Scale	1. Spivack & Spotts, 1967	640 emotionally disturbed, mentally retarded, and normal adolescents, aged 13 to 18 years. IQ ranges not reported.	"Factor reliability" ranged from .57 (Anxious Self-Blame) to .86 (Unethical Behavior, & Dominating-Sadistic). Mean = .77.	

Test-Retest	Interrater	VALIDITY: Factorial/ Taxonomic	Criterion Group	Congruent
		Cluster analysis rendered eight clusters characterized as (1) essentially normal (2 clusters), (2) hyperkinetic syndrome, (3) stereotypy/emotional withdrawal, (4) high arousal with multiple disturbances, (5) affective-like disorders, (6) pathological social withdrawal, and (7) withdrawal characteristic of dementia.		
Subjects followed up over 6 years showed no overall change in severity of psychiatric disorder. Correlations (tau $_B$) for manifest symptoms ranged from .12 to .58 (M=.38). 11 of 13 manifast abnormalities consistent over 6 years; for 5 of 13 symptoms tau $_B$ >.50				
	Reliability of reported symptoms = .78; reliability of manifest abnormalities = .85. Sample size = 5 subjects; reliabilities for individual symptoms not reported.	Factor analysis (principal components with varimax rotation) of CIS ratings resulted in eight-factor solution: (1) Neurotic Depression (2) Neurasthenia, (3) Mental Retardation, (4) Psychoticism, (5) Medication Effects, (6) Phobias, (7) Elation, and (8) Hypochondria. Cluster analysis rendered seven clusters, the largest of which (65%) reflected no disturbance.		Factors derived from CIS correlated with Behaviour Disturbance Scale as follows: Communicativeness on BDS significantly related to Neurasthenia and Retardation. Aggression on BDS associated with Phobia on CIS. Authors concluded that the psychiatric (CIS) and behavioral indices were not strongly related.
		Factor analysis of 125 items resulted in an 18-factor solution. Correlational analysis of 47 additional items resulted in four item clusters over and above the 18 factors.		

Instrument	Authors	Samples	RELIABILITY: Internal Consistency (alpha)	Item Total Correlations
(DAB cont.)	2. Spivack, Haimes, & Spotts, 1967	a) 315 disturbed adolescents residing in three institutions b) 141 mentally retarded adolescents residing in an institution c) 92 normal adolescents residing in an institution d) 305 normal adolescents residing at home.		
Devereux Child Behavior Rating Scale	1. Spivack & Levine, 1964 (preliminary version of scale)	140 institutionalized children, aged 5 to 12 years. IQs ranged from 30 to over 100; 59% of sample had IQs less than 80.		
	2. Spivack & Spotts, 1965	252 children, aged 6 to 12 years, residing in four institutions. IQ ranged from less than 20 to over 100, with a mean of 71.		
	3. Spivack & Spotts, 1966	a) Same sample as in Spivack & Spotts (1965), above. b) 100 mentally retarded children, aged 6 to 13 years, and with IQs ranging from less than 20 to over 100 (92% had IQs < 60). c) 348 public school children, presumed normal.		

Test-Retest	Interrater	VALIDITY: Factorial/ Taxonomic	Criterion Group	Congruent
7-to-10 day test-retest reliability for a mixed treatment sample of 83 adolescents ranged from .53 (Hyperactivity) to .91 (Schizoid With- . drawal) Mean=.81.	a) Sample of 89 disturbed adolescents: Correlations ranged from .01 (Anxious Self-Blame) to .68 (Bizarre Action). Mean = .40. b) Sample of 254 normal adolescents: Correlations ranged from .22 (Need Approval) to .66 (Heterosexual Interest). Mean=.43.	12 factors and 3 clusters loosely modeled after results of Spivack & Spotts (1967), above.	Mean subscale scores for disturbed clinical groups (by diagnosis), mentally retarded adolescents, and normal adolescents found to differ for most subscales.	
	Intraclass correlation coefficients ranged from .77 (Receptor Hyposensitivity subscale) to .93 (Arrested Self-Care) across factors. Mean intraclass correlation coefficient was .84.	Factor analysis of 68-item instrument rendered 15 factors, many of which were similar to factors on final scale.		
		Factor analysis of 121-item instrument rendered 20 factors. Six second-order factors also derived.		
One-week test-retest data: Correlations ranged from .80 to .99 across subscales (M=.90). One-month data: Correlations ranged from .77 to .96 (M=.85). 6-month data: Correlations ranged from .35 to .75 (M=.60). Sample sizes not reported.			Large majority of subscale scores appeared to be lower for normal children as compared with children having behavioral/emotional disorders or mental retardation. (Inferential statistics showing significance not reported.)	

Instrument	Authors	Samples	RELIABILITY: Internal Consistency (alpha)	Item Total Correlations
Diagnostic Assessment of the Severely Handicapped (DASH) Scale	1. Matson, Coe, Gardner, & Sovner, 1989	506 severely and profoundly mentally retarded residents (including 247 females and 254 males, mean age 38 years) of four developmental centers.	Alpha coefficients for 6 derived factors ranged from .62 to .80 (M=.69).	
	2. Matson, Gardner, Coe, & Sovner, 1989b.	Same as #1, above.	Alpha ranged from .20 (Schizophrenia) to .84 (M=.53).	
Emotional Disorders Rating Scale for Developmental Disabilities (EDRS-DD)	1. Feinstein, Kaminer, Barrett, & Tylenda, 1988	10 psychiatrically disordered children and adolescents, aged 9 to 20 years, in a developmental disabilities unit of a children's hospital.		
	2. Kaminer, Seifer, Stevens, & Barrett, in press.	39 patients, aged 7 to 17 years, with IQ ≥ 85. 13 subjects were in-patients, whereas 26 were day patients. Subjects not mentally retarded.	Alpha coefficients ranged from .00 (Somatic/Vegetative) to .86 (Hostility/Anger). Mean alpha=.51.	
Minnesota Developmental Programming System: Behavior Management Assessment (BMA)	1. Olvera, Bock, & Silverstein, 1985	Mean Behavior Management scores presented separately for each of 10 types of residential settings in Indiana. No *SD*s reported.		

Test-Retest	Interrater	VALIDITY: Factorial/ Taxonomic	Criterion Group	Congruent
		Factor analysis produced a 6-factor solution encompassing 41 items and explaining 39% of the variance. Factors were labeled as: (1) Emotional Lability, (2) Aggression/ Conduct Disorder, (3) Language Disorder/ Verbal Aggression, (4) Social Withdrawal, (5) Eating Disorder, & (6) Sleep Disorder.		
	Agreement levels for 29 pairs of interviewers and informants (calculated by dividing agreement by agreement-plus-disagreement) were as follows: (1) Frequency, .91; (2) Duration, .95; (3) Severity, .96.	Subscale composition largely guided by structure of DSM–III–R.		
	Correspondence of ratings over 5 days said to range from 85% to 96% for frequency ratings and from 86% to 96% for severity ratings. (No account taken of chance levels of agreement.)	Items derived from DSM-III criteria for major affective disorders, observable anxiety symptomatology, and from clinical experience.		
Correspondence over 1 week ranged from -.14 (Depressed Mood-Verbal) to .84 (Hostility/Anger). Mean=.39.	Reliability ranged from .62 (Irritability) to .82 (Hostility/ Anger). Mean=.72. Kappa statistic reported in general terms (mostly modest) but not summarized for specific items.		Patients diagnosed as depressed rated significantly higher on Non-Verbal Depression and significantly lower on Manic/Elated Mood.	Correspondences betwen EDRS (Depressed Mood–Verbal items) and Children's Depression Rating Scale and the Hamilton Depression Rating Scale were .63 and .72, respectively.

Instrument	Authors	Samples	RELIABILITY:	
			Internal Consistency (alpha)	Item Total Correlations
(MDPS-BMA cont.)	2. Silverstein, Olvera, & Schalock, 1989			
Preschool Behavior Questionnaire	1. Hammer, Kimball, & Beck, 1989	20 preschool boys of normal IQ; 20 boys in preschools for children with developmental delays. 10 subjects in each group had Attention Deficit Disorder with Hyperactivity.		
	2. Rheinscheld, 1989	203 children attending Early Childhood Education Centers operated by County Boards of Mental Retardation and Developmental Disabilities.		

Test-Retest	Interrater	VALIDITY: Factorial/ Taxonomic	Criterion Group	Congruent
	Reliability for two sets of 40 behavior technicians = .66 for intensity scores.			Correlations between intensity scores on BMA and staff time for behavioral habilitation = .67 to.81 (M=.76). Mean correlation with staff time per program unit = .86.
				Hyperactive-Distractible subscale correlated with teacher judgments of hyperactivity using DSM-III criteria (R^2=.40). Hyperactive-Distractible ratings correlated with commission errors (.41) on Continuous Performance Task but not omission errors or observations of playroom behavior.
		Factor analysis of ratings using 3-factor solution corresponded closely with original structure reported by Behar and Stringfield (1974). 21 of 24 items (88%) loaded most heavily on same respective subscales. Model accounted for 46% of variance.		Global ratings of activity level correlated with Hostile-Aggressive (.33) and Hyperactive-Distractible (.32) subscales.

Instrument	Authors	Samples	RELIABILITY: Internal Consistency (alpha)	Item Total Correlations
Prout-Strohmer Personality Inventory (PSPI)	1. Prout & Strohmer, 1989	708 adolescents and adults with mild mental retardation or borderline intelligence, placed in a variety of day programs and residential programs.	Alpha coefficients ranged from .77 to .89 across clinical scales and lie scale (mean=.84).	These ranged from .20 to .66 between items and respective subscales. Mean correlations ranged from .38 (Depression) to .51 (Low Self-Esteem) Overall mean = .40.
Psychopathology Instrument for Mentally Retarded Adults (PIMRA)	1. Kazdin, Matson, & Senatore, 1983	No normative sample. Validation sample (N= 110) of mentally retarded adults aged 18-71 years, 74 of whom (67%) had psychiatric diagnoses of mental disorder. Level of mental retardation ranged from borderline to severe.		

Test-Retest	Interrater	VALIDITY:		
		Factorial/ Taxonomic	Criterion Group	Congruent
Two studies completed. With 4- to 6-week test-retest interval, correlations ranged from .65 (Low Self-Esteem) to .89 (Thought/Behavior Disorder). Mean=.81. With 2-week test-retest interval, correlations ranged from .66 (Low Self-Esteem) to .85 (Depression, Thought Disorder). Mean=.80.	NA	Items compiled and assigned to subscales using "rational/clincial" approach. Item placement supported by confirmatory factor analysis. Subscales highly intercorrelated: Range = .52 to .76 (M=.64).	Clincial subscale scores N.S. higher for each of the following index groups: (1) Subjects taking psychotropic drugs. (2) Subjects having behavior plan to reduce problem behaviors. (3) Subjects having a DSM-III diagnosis indicating externalizing behavior problem. (4) Subjects regarded as emotionally disturbed or psychotic. (5) Subjects in day treatment programs. (6) Subjects in restrictive residential programs. Inferential statistics not presented.	Correlations between PSPI Anxiety subscale and four scores from the Children's Manifest Anxiety Scale ranged from .76 to .88. Correlation between PSPI Depression subscale and Beck Depression Inventory = .74. Correlations between PSPI subscales and respective dimensions from Strohmer-Prout Behavior Rating Scale ranged from .13 to .20 (mean =.17). Low to moderate correlations between PSPI scores and counselor global ratings of emotional adjustment.
				Self-report scores on Depression subscale correlated with Beck depression ratings (r=.33), but not with Zung, Thematic Apperception Test, or MMPI Depression scores. PIMRA Total self-report scores significantly correlated with Beck, Zung, and MMPI but not Thematic Apperception Test Depression Ssores. Informant Depression scores correlated significantly (r=.74) with Hamilton Depression scores.

Instrument	Authors	Samples	RELIABILITY: Internal Consistency (alpha)	Item Total Correlations
(PIMRA cont.)	2. Matson, Kazdin, & Senatore, 1984a	Same as #1, above.		
	3. Matson, Kazdin, & Senatore, 1984b	Same as #1, above.		
	4. Senatore, Matson, & Kazdin, 1985	Same as #1, above.	(a) Self-report version alpha=.85 Spearman-Brown split-half = .88. (b) Informant version alpha=.83 Spearman-Brown split-half = .65 Computed for Total Score only; no data presented for individual subscales.	(a) Self-report version - mean r=.35 - 46 of 56 items (82%) correlated significantly with the Total Score. (b) Informant version - mean r=.35 - 41 of 56 items (73%) correlated with Total Score.

Test-Retest	Interrater	VALIDITY: Factorial/ Taxonomic	Criterion Group	Congruent
			On self-report version, no differences were found between medicated and unmedicated subjects. On informant version, subjects taking medication (primarily psychotropic) had higher scores on Schizophrenia, Affective Disorder, and Adjustment Disorder. Dose-related findings also reported.	
		Factor analysis of self-report ratings rendered two factors labeled (1) Anxiety and (2) Social Adjustment. Factor analysis of informant ratings rendered three factors: (1) Affective Disorder, (2) Somatoform Disorder, and (3) Psychosis.		
a) Self-report ratings: correlations ranged from .42 to .69. Four of 8 subscales below .60. (b) Informant ratings: correlations ranged from .48 to 1.00 (M=.76). One of 8 subscales correlated below .60.	Informant vs. Self-report ratings: only 10 of 56 items (18%) were significantly correlated.	Items adapted from DSM-III.	Subjects with diagnosed psychopathology had higher Total Scores on informant version than subjects with no diagnoses. No data on specific diagnoses reported.	

Instrument	Authors	Samples	RELIABILITY:	
			Internal Consistency (alpha)	Item Total Correlations
(PIMRA cont.)	5. Aman, Watson, Singh, Turbott, & Wilsher, 1986; Watson, Aman, & Singh, 1988	Sample made up of 95 adults attending a workshop training center and 65 adults residing in a developmental center. Mental retardation ranged from borderline to severe.	Coefficient alpha ranged from .45 to .73 on the self-report version (M=.64). On informant version, alpha ranged from .60 to .71 (M=.66)	Mean item-total correlation for self-report version =.40. Excluding Personality subscale, 82% of correlations were significant. For informant version, mean correlation =.46. Excluding Personality subscale items, 93% of correlations were significant.
	6. Helsel & Matson, 1988	Sample for psychometric purposes comprised of 99 adults with mental retardation, aged 17 to 57 years, and with level of mental retardation ranging from borderline to severe.		
	7. Davidson, 1988 (in Matson, 1988)	244 adults in community-based or residential programs.		
	8. Tymchuk, 1989	31 mothers with mild mental retardation; 97 mothers of normal IQ.	Self-report version: Coefficient alpha ranged from .06 to .56 (M=.32).	Self-report version: Correlations ranged from .30 (Psychosexual Disorder) to .52 (Adjustment Disorder) (M=.44).

Test-Retest	Interrater	VALIDITY: Factorial/ Taxonomic	Criterion Group	Congruent
Only 3 of 8 subscale scores were significantly correlated on self-report version over 5 months. Range = -.15 to .56; M=.31).	Informant vs. self-report ratings showed significant relationships for only 4 of 8 subscales (range = -.05 to .58; mean correlation =.19).	Factor analysis of self-report and informant ratings resulted in four factors each. The factors were labeled as fol-follows: Self-report: (1) Anxiety, (2) Social Adjustment, (3) Identity/Reality Concern, (4) Unlabeled (mixed). Informant: (1) Affective Concerns, (2) Social Adjustment, (3) Somatoform Difficulty, and (4) Unlabeled.	Moderately retarded subjects had significantly higher scores on Schizophrenia than did subjects with mild mental retardation. Mildly retarded subjects scored significantly higher on Affective Disorder than did subjects with moderate retardation. No differences were found between subjects residing in a developmental center and those living in the community.	
				PIMRA self-report Depression scale correlated with Beck ratings but not Zung or Hamilton Depression ratings. PIMRA informant ratings of Depression not correlated with self ratings of depression on Beck, Zung, Hamilton, or PIMRA scales. PIMRA Total Scores (self & informant) correlated with scales measuring depression in 3 of 6 comparisons.
				Total Score correlated .83 with CHEMRA, an antecedent of the Reiss Screen for Maladaptive Behavior.
			Mothers with mental retardation had significantly higher scores than mothers of normal IQ on all except the Psychosexual Disorder subscale.	

Instrument	**Authors**	**Samples**	**RELIABILITY: Internal Consistency (alpha)**	**Item Total Correlations**
(PIMRA cont.)	9. Iverson & Fox, 1989	Random sample of 165 adults, stratified for level of mental retardation (mild to profound) and living environment (institutional vs. family vs. independent). 36% of sample found to have at least one significant psychopathological disorder.		
	10. Sturmey & Ley, 1990		Informant version: Coefficient alpha ranged from.04 to .69 (M=.41).	Informant version: Point biserial correlations ranged from -.32 to .77 with median of .29. 5 of 8 subscales had median correlations below .30.
Reiss Screen for Maladaptive Behavior	1. Reiss, 1988b	Normative sample N=258 Validation sample N=418	(a) alpha = .54 to .84 (M=.74) (b) alpha = .57 to .85 (M=.73).	

Test-Retest	Interrater	VALIDITY: Factorial/ Taxonomic	Criterion Group	Congruent
	Percentage agreement ranged from 70% (Anxiety subscale) to 95% (Psychosexual subscale) with a mean agreement of 80%. (No correction made for chance level of agreement). Agreement occurred on 17 of 19 subjects (89%) for presence of significant psychopathology.			
				Several subscales from the PIMRA correlated with subscales on the Aberrant Behavior Checklist (ABC) (Aman & Singh, 1986) as follows: Schizophrenic& Lethargy, withdrawal (ABC); Affective & Lethargy, withdrawal; Adjustment & Hyperactivity; Personality & Lethargy, withdrawal/Stereotypic behavior; Inapprop. mental adjustment & Stereotypy. Median r=.62.
	Item by item only: .30 - .73, mean=.54 (generally high).	First seven scales factor analytically derived; factor loadings generally high (M=.59).Factor structure said to be validated for Spanish version of Reiss Screen (Sacristan, 1987; cited in Reiss, 1988).	Mixed group of subjects with dual diagnosis (n=112) had significantly higher scores than those with no diagnosis (n=167). Reiss Screen correctly classified 43 of 59 subjects (73%) who received full diagnostic work-up.	Total Score on antecedent of Reiss Scale correlated highly with total PIMRA score (r=.83) (Davidson, 1988; cited in Reiss, 1988). Correlations of Total Score with AAMD Part II also high (r=.78).

Instrument	Authors	Samples	RELIABILITY: Internal Consistency (alpha)	Item Total Correlations
Schedule of Handicaps, Behaviour, & Skills (HBS)	1. Wing, 1978	84 "psychotic" children (having autistic traits), aged 2 to 18 years. 74 children, under 15 years of age, with severe mental retardation and who were not socially aloof.		
	2. Wing & Gould, 1978	104 children, aged 2 to 18 years, receiving services for mental retardation. Approximately 85% of sample had IQs below 50.		

Test-Retest	Interrater	VALIDITY: Factorial/ Taxonomic	Criterion Group	Congruent
			The "psychotic" subjects differed from the remainder on the following variables: (1) Lack of eye contact, (2) Presence of marked stereotypies, (3) Presence of elaborate routines, (4) Expression of symbolic play, and (5) Lack of sociability. Both classification and differences appear to be based on HBS Schedule	
	Between diagnosers using audiotapes of interviews: Agreement occurred for all 20 subjects studied, on all except "Repetitive symbolic play" section, where agreement occurred for 19 of 20 subjects. Between types of informants (parents vs. professionals): 3 indices of agreement, (Maximum Aggreement [MA], Agreement for Presence [AP], & Agreement for Absence [AA] of symptoms [see text]) were used. 80% or more of subjects were correctly classified on 7/20, 2/20, & 7/20 sections using MA, AP, AA, respectively. Fewer than 70% of subjects were classified on 4/20, 16/20, & 4/20 sections using MA, AP, & AA, respectively.			

Instrument	Authors	Samples	RELIABILITY: Internal Consistency (alpha)	Item Total Correlations
(HBS cont.)	3. Wing & Gould, 1979 (also, Wing, 1975; Wing, 1981)	132 children, aged 2 to 18 years, moderate to profound mental retardation, residing in a London borough.		
	4. Bernsen, 1980	148 children, aged 3 to 22 years, with IQs less than 50.		
	5. Lund, 1985	All relevant adults (N=302) living in part of a Danish county, selected to be representative of the Danish population. HBS Schedule, supplemented by a list of psychiatric symptoms, was used to assess all subjects for psychiatric disorder.		
Self-Report Depression Questionnaire	1. Reynolds & Baker, 1988	83 adults, aged 21 to 72 years, with IQs ranging from 35 to 75.	Coefficient alpha equaled .90 and .93 over two administrations.	These ranged from .27 to .68, with a mean of .45.

Test-Retest	Interrater	VALIDITY: Factorial/ Taxonomic	Criterion Group	Congruent
			Children were divided into "Socially Impaired" (aloof) (N=74) and "Sociable" (N=58) groups. Socially Impaired subjects had higher levels than the Sociable children on each of the following: muteness or echolalia, lack of symbolic activities, language comprehension, organic conditions, age of onset after birth, and a disproportionate number of males.	
	Mean agreement between parents and professional informants = .70, .66, & .43 using MA, AP, & AA (see above), respectively.			
Over an 11-week interval, reliability was .63 for 44 subjects.	NA	Scale items related to symptoms of Depression and Dysthymic Disorder on DSM-III-R, Research Diagnostic Criteria for depression, and Hamilton (1960) Depression Rating Scale. Exploratory factor analysis produced a 10-factor solution accounting for 68% of the variance.		SRDQ scores were correlated with interview scores on Hamilton Depression Rating Scale: correlations of .65 and .63 were obtained at two assessment times.

Instrument	Authors	Samples	RELIABILITY:	
			Internal Consistency (alpha)	Item Total Correlations
Strohmer-Prout Behavior Rating Scale (SPBRS)	1. Strohmer & Prout, 1989	673 adolescents and adults, with borderline IQs or mild mental retardation, placed in a wide variety of day and residential programs.	Alpha coefficients ranged from .90 to .96 (mean=.93) across subscales.	These ranged from .30 to .89 between items and respective subscales. Mean correlations ranged from .62 (Thought/ Behavior Disorder) to .78 (Somatic Concerns). Overall mean = .71.
Vineland Adaptive Behavior Scale	1. Sparrow, Balla, & Cicchetti, 1984	a) 3,000 subjects, stratified by sex, race, community size, region, and parental education. b) Supplementary groups made up of 1,150 mentally retarded, 150 emotionally disturbed, 200 visually handicapped, and 300 hearing-impaired subjects.	Spearman-Brown split-half reliability ranged from .77 to .88 (mean=.85) across ages.	
	2. Volkmar, Sparrow, Goudreau, Cicchetti, Paul, & Cohen, 1987	a) 35 children with autism or Childhood Onset Pervasive Developmental Disorder. Mean age= 12.4 years. b) 22 nonautistic children with developmental disabilities. Mean age=11.1 years.		

Test-Retest	Interrater	VALIDITY: Factorial/ Taxonomic	Criterion Group	Congruent
	Assessed on 42 subjects by raters in same setting: Correlations ranged from .24 (Sexual Maladjustment) to .95 (Physical Aggression). Mean = .82. Assessed on 26 subjects by raters in different settings: Correlations ranged from .44 (Sexual Maladjustment) to .93 (Physical Aggression). Mean=.78.	Items compiled and assigned to subscales using "rational/ clinical" approach. Item placement confirmed by item-subscale correlations and confirmatory factor analysis. Subscales mildly to moderately intercorrelated: Range = .09 to .80; mean correlation = .41.	Subscale scores generally higher for each of the following index groups: (1) Subjects taking psychotropic drugs. (2) Subjects having plan to reduce problem behaviors. (3) Subjects having a DSM-III diagnosis indicating externalizing behavior problem. (4) Subjects regarded as emotionally disturbed or psychotic. (5) Subjects in day treatment programs. (6) Subjects in restrictive residential placements. (Inferential statistics not presented.)	1. Subscale scores on SPBRS moderately to strongly correlated with analogous subscales on Child Behavior Checklist. 2. Subscale scores on SPBRS moderately to strongly correlated with analogous maladaptive scales on AAMD Adaptive Behavior Scale and Inventory for Client and . Agency Planning (ICAP). 3. SPBRS subscale scores weakly correlated with self-rating scores on the Prout-Strohmer Personality Inventory. 4. SPBRS subscale scores moderately correlated with global behavior ratings by counselors.
2-to-4 week reliability ranged from .84 to .89 across ages. Mean=.88.	For 94 subjects, interrater reliability was .74.		1. Means for supplementary groups higher than for national standardization sample on Part 1. 2. The emotionally disturbed sample obtained higher scores than the other supplementary groups, both on Part 1 and Part 2.	
			Autistic subjects received significantly higher Part 1 and Part 2 scores than nonautistic developmentally delayed subjects.	

Appendix C

Full Instrument Names for Abbreviations Used in Tables 3 Through 7

Abbreviation	Instrument Name
ABS:R	AAMD Adaptive Behavior Scale: Residential and Community Edition
ABS:S	AAMD Adaptive Behavior Scale: School Edition
ABC	Aberrant Behavior Checklist
Adolesc. Behav. CL	Adolescent Behavior Checklist
Attention CL	Attention Checklist
BDS$_1$	Behaviour Disturbance Scale
BDS$_2$	Behavior Development Survey
BeERS	Behavior Evaluation Rating Scale
BIRD	Behavior Inventory for Rating Development
BPI	Behavior Problems Inventory
BSAB-II	Balthazar Scales of Adaptive Behavior: II. Scales of Social Adaptation
CIS	Clinical Interview Schedule
CDER	Client Development Evaluation Report
Comm. Style Q	Communication Style Questionnaire
DABRS	Devereux Adolescent Behavior Rating Scale
DASH	Diagnostic Asssessment of the Severely Handicapped Scale
DCBRS	Devereux Child Behavior Rating Scale
DDCBCL	Developmentally Delayed Children's Behavior Checklist
EDRS–DD	Emotional Disorders Rating Scale – Developmental Disabilities

Fairview	Fairview Maladaptive Behavior Survey
Gilson-Levitas	Gilson-Levitas Diagnostic Criteria
HBS	Schedule of Handicaps, Behaviour, and Skills
MAS	Motivation Assessment Scale
MDPS-BMA	Minnesota Developmental Programming System: Behavior Management Assessment
PBQ	Preschool Behavior Questionnaire
PBS	Psychosocial Behaviour Scale
PIMRA	Psychopathology Instrument for Mentally Retarded Adults
PSPI	Prout-Strohmer Personality Inventory
RCMAS	Revised Children's Manifest Anxiety Scale
Reiss	Reiss Screen for Maladaptive Behavior
SAP	Standardized Assessment of Personality
SCI	Structured Clinical Interview
SEBI	Social and Emotional Behavior Inventory
SJS	Social Judgment Scale
SPBRS	Strohmer-Prout Behavior Rating Scale
Soc. Part. RS	Social Participation Rating Scale
SRDQ	Self-Report Depression Questionnaire
Vineland	Vineland Adaptive Behavior Scales
Zung	Zung Self-Rating Anxiety Scale (Adapted)

☆ U.S. GOVERNMENT PRINTING OFFICE: 1991 — 281-820/44105

DHHS Publication No. (ADM) 91-1712
Alcohol, Drug Abuse, and Mental Health Administration
Printed 1991